1 Proof of main theorem

The main result of this section is

**Theorem 1.** There exists an absolute constant $C > 0$ such that for every $\delta \in (0, 1)$, every integer $1 \leq m \leq n^4$ and every matrix $U \in \mathbb{R}^{n \times d}$ with orthonormal columns, if $B \geq \frac{1}{2} C (\log n)^4 \cdot d^2 \cdot m^{1/2}$, $S \in \mathbb{R}^{B \times n}$ is a random CountSketch matrix, and $G \in \mathbb{R}^{m \times B}$ and $\tilde{G} \in \mathbb{R}^{m \times n}$ are matrices of i.i.d. unit variance Gaussians, then the total variation distance between the joint distribution $GSU$ and $\tilde{G}U$ is less than $\delta$.

**Remark 2.** Note that we restrict the range of values of $m$ in Theorem 1 to $[1 : n^4]$. This is because if $m > n^4$, the theorem requires $B \gg \frac{1}{3} n^2$, at which point the CountSketch matrix $S$ becomes an isometry of $\mathbb{R}^n$ with high probability and the theorem follows immediately. At the same time restricting $m$ to be bounded by a small polynomial of $n$ simplifies the proof of Theorem 1 notationally.

Recall that a CountSketch matrix $S \in \mathbb{R}^{B \times n}$ is a matrix all of whose columns have exactly one nonzero element in a random location, and the value of the nonzero element is independently chosen to be $-1$ or $+1$. All random choices are made independently. Throughout this section we denote the number of rows in the CountSketch matrix by $B$. Note that the matrix $S$ is a random variable. Let $G$ denote an $m \times B$ matrix of independent Gaussians. For an $n \times d$ matrix $U$ with orthonormal columns let $q : \mathbb{R}^d \to \mathbb{R}_+$ denote the p.d.f. of the random variable $G_1SU$, where $G_1$ is the first row of $G$ (all rows have the same distribution and are independent). We note that $G_1SU$ is a mixture of Gaussians. Indeed, for any fixed $S$ the distribution of $G_1SU$ is normal with covariance matrix $(G_1SU)^T(G_1SU) = U^T S^T SU$. We denote the distribution of $G_1SU$ given $S$ by

$$q_S(x) := \frac{1}{\sqrt{(2\pi)^d \det M}} e^{-\frac{1}{2} x^T M^{-1} x}.$$  

Throughout this section we use the notation $M := U^T S^T SU$. Note that since $S$ is a random variable, $M$ is as well. With this notation in place we have for any $x \in \mathbb{R}^d$

$$q(x) = \mathbb{E}_S[q_S(x)].$$  

Let $p : \mathbb{R}^d \to \mathbb{R}_+$ denote the pdf of the isotropic Gaussian distribution, i.e. for all $x \in \mathbb{R}^d$

$$p(x) = \frac{1}{\sqrt{(2\pi)^d}} e^{-\frac{1}{2} x^T x}. $$

Before giving a proof of Theorem 1, which is somewhat involved, we give a simple proof of a weaker version of the theorem, where the number of buckets $B$ of our CountSketch matrix is required to be $\approx \frac{1}{3} d^2 m$ as opposed to $\approx \frac{1}{6} d^2 \sqrt{m}$:

**Theorem 3.** There exists an absolute constant $C > 0$ such that for every $\delta \in (0, 1)$, every integer $m \geq 1$ and every matrix $U \in \mathbb{R}^{n \times d}$ with orthonormal columns if $B \geq \frac{1}{2} C d^2 \cdot m$, $S \in \mathbb{R}^{B \times n}$ is a random CountSketch matrix, and $G \in \mathbb{R}^{m \times B}$ and $\tilde{G} \in \mathbb{R}^{m \times n}$ are matrices of i.i.d. unit variance Gaussians, then the total variation distance between the joint distribution $GSU$ and $\tilde{G}U$ is less than $\delta$.

We will use the following measures of distance between two distributions in the proof of our main theorem (Theorem 1) as well as the proof of Theorem 3:

**Definition 4** (Kullback-Leibler divergence). The Kullback-Leibler (KL) divergence between two random variables $P, Q$ with probability density functions $p(x), q(x) \in \mathbb{R}^d$ is given by $D_{KL}(P||Q) = \int_{\mathbb{R}^d} p(x) \ln \frac{p(x)}{q(x)} dx$.

**Definition 5** (Total variation distance). The total variation distance between two random variables $P, Q$ with probability density functions $p(x), q(x) \in \mathbb{R}^d$ is given by $D_{TV}(P, Q) = \frac{1}{2} \int_{\mathbb{R}^d} |p(x) - q(x)| dx$.

**Theorem 6** (Pinsker’s inequality). For any two random variables $P, Q$ with probability density functions $p(x), q(x) \in \mathbb{R}^d$ one has $D_{TV}(P, Q) \leq \sqrt{\frac{1}{2} D_{KL}(P||Q)}$.  

1
The proof of Theorem 3 uses the following simple claim.

Claim 7 (KL divergence between multivariate Gaussians). Let $X \sim N(0, I_d)$ and $Y \sim N(0, \Sigma)$. Then $D_{KL}(X||Y) = \frac{1}{2} \text{Tr}(\Sigma^{-1} - I) + \frac{1}{2} \ln \det \Sigma$.

Proof. One has

$$D_{KL}(X||Y) = E_{X \sim N(0, I_d)}[\frac{1}{2} X^T X + \frac{1}{2} X^T \Sigma^{-1} X + \frac{1}{2} \ln \det \Sigma]$$

$$= E_{X \sim N(0, I_d)}[\frac{1}{2} X^T (\Sigma^{-1} - I) X + \frac{1}{2} \ln \det \Sigma]$$

$$= \frac{1}{2} E_{X \sim N(0, I_d)}[\text{Tr}((\Sigma^{-1} - I) XX^T)] + \frac{1}{2} \ln \det \Sigma$$

$$= \frac{1}{2} \text{Tr}(\Sigma^{-1} - I) + \frac{1}{2} \ln \det \Sigma,$$

where we used the fact that for a vector $X$ of independent Gaussians of unit variance one has $E_X[X^TAX] = \text{Tr}(A)$ for any symmetric $A$ (by rotational invariance of the Gaussian distribution).

We can now give

Proof of Theorem 3. One has by Lemma 21 (1) (see below; this is a standard property of the CountSketch matrix) that for any $U \in \mathbb{R}^{n \times d}$ with orthonormal columns, and $B \geq 1$, if $S$ is a random CountSketch matrix and $M = U^T S^T S U$, then $E_S[||M - I||_F^2] = O(d^2/B)$. By Markov’s inequality $Pr_S[||I - M||_F > (2/\delta) \cdot O(d^2/B)] < \delta/2$. Let $E$ denote the event that $||I - M||_F \leq (2/\delta) \cdot O(d^2/B)$. We condition on $E$ in what follows. Since $B \geq \frac{2}{\delta} Cd^2m$ for a sufficiently large absolute constant $C > 1$, we have, conditioned on $E$, that

$$||I - M||_F^2 \leq (2/\delta) \cdot O(d^2/B) = (2/\delta) \cdot \frac{\delta^3}{(Cm)} \leq 2\delta^2/(Cm).$$

Note that in particular we have $||I - M|| \leq ||I - M||_F < 1/2$ conditioned on $E$ as long as $C > 1$ is larger than an absolute constant.

By Claim 7 we have $D_{KL}(X||Y) = \frac{1}{2} \text{Tr}(I - \Sigma^{-1}) + \frac{1}{2} \ln \det \Sigma$. We now use Taylor expansions of matrix inverse and log det provided by Claim 9 and Claim 10 (see below) to obtain

$$D_{KL}(X||Y) = \frac{1}{2} \text{Tr}(M^{-1} - I) + \frac{1}{2} \ln \det M$$

$$= \frac{1}{2} \text{Tr}\left( \sum_{k \geq 1} (I - M)^k \right) + \frac{1}{2} \sum_{k \geq 1} \frac{(-\text{Tr}(I - M)^k)}{k}$$

$$= \frac{1}{2} \text{Tr}\left( \sum_{k \geq 2} (I - M)^k \right) + \frac{1}{2} \sum_{k \geq 2} \frac{(-\text{Tr}(I - M)^k)}{k}$$

$$= O(\text{Tr}((I - M)^2))$$

(as long as $C > 1$ is larger than an absolute constant. This shows that for every $S \in E$ one has $D_{KL}(p||q_S) \leq (\delta/4)^2/m$, and thus $D_{KL}(p||q_E) \leq (\delta/4)^2/m$, where we let $q(x) := E_S[q_S(x)|E]$.

We now observe that the vectors $(G_iSU)_{i=1}^m$ and $(G_iU)_{i=1}^m$ are vectors of independent samples from distributions $q(x)$ and $p(x)$ respectively. We denote the corresponding product distributions by $q^m$ and $p^m$. Since the good event $E$ constructed above occurs with probability at least $1 - \delta/2$, it suffices to consider the distributions $\tilde{q}(x)$ and $p(x)$, as

$$D_{TV}(q^m, p^m) \leq Pr[E] + D_{TV}(q^m, p^m|E) = Pr[E] + D_{TV}((\tilde{q})^m, p^m),$$

(5)
where $D_{TV}(q^m, p^m|\mathcal{E}) = D_{TV}(\tilde{q}^m, p^m)$ stands for the total variation distance between the distribution of $(G_iU)_i^{m}$ and the distribution of $(G_iSU)_i^{m}$ conditioned on $S \in \mathcal{E}$. We can now use the estimate from (4) to get

$$
D_{TV}(\tilde{q}^m, p^m) \leq \sqrt{\frac{1}{2} D_{KL}(p^m||\tilde{q}^m)} \text{ (by Pinsker’s inequality)}
$$

$$
= \sqrt{\frac{m}{2} D_{KL}(p||\tilde{q})} \text{ (by additivity of KL divergence over product spaces)}
$$

$$
\leq \sqrt{\frac{m}{2} \cdot (\delta/4)^2/m} \text{ (by (4))}
$$

$$
\leq \delta/4.
$$

The main source of hardness in proving the stronger result provided by Theorem 1 comes from the fact that unlike the setting of Theorem 3, where most elements in the mixture are close to isotropic Gaussians in KL divergence, in the setting of Theorem 1 most elements of the mixture are too far from isotropic Gaussians to establish our result directly (this can be seen by verifying that the bounds of Theorem 3 on the KL divergence of $q_S$ to $p$ are essentially tight). Thus, the main technical challenge in proving Theorem 1 consists of analyzing the effect of averaging over random CountSketch matrices that is involved in the definition of $q(x)$ in (1). The core technical result behind the proof of Theorem 1 is Lemma 8, stated below.

Ideally, we would like a lemma that states that the ratio of the pdfs $q(x)/p(x)$ is very close to 1 for ‘typical’ values of $x$ (for appropriate definition of a set of ‘typical’ $x$). Unfortunately, it is not clear how to achieve this result for the distribution $q(x)$ defined in (1). The problem is that some choices of CountSketch matrices $S$ may lead to degenerate Gaussian distributions that are hard to analyze. For example, when $S$ is not a subspace embedding, the matrix $M$ may even be rank-deficient, and the inverse $M^{-1}$ is then ill-defined. To avoid these issues, we work with an alternative definition. Specifically, instead of averaging the distributions $\frac{1}{\sqrt{(2\pi)^d \det M}} e^{-\frac{1}{2} x^T M^{-1} x}$ over all CountSketch matrices, we define a high probability event $\mathcal{E}$ in the space of matrices $S$ (see Lemma 8 for the definition) and reason about the modified distribution $\tilde{q}(x)$ defined as

$$
\tilde{q}(x) = E_S \left[ \frac{1}{\sqrt{(2\pi)^d \det M}} e^{-\frac{1}{2} x^T M^{-1} x} \right] \mathcal{E}.
$$

For technical reasons it turns out to be useful to define yet another distribution

$$
q'(x) = E_S \left[ \frac{1}{\sqrt{(2\pi)^d \det M}} e^{-\frac{1}{2} x^T M^{-1} x} \cdot I[x \in \mathcal{T}(S,U)] \right] \mathcal{E} + \xi \cdot p(x),
$$

where $\xi = E_S [Pr_{X \sim q_S} [X \notin \mathcal{T}(S,U)] ] \mathcal{E} \leq n^{-20}$ and for each $S \in \mathcal{E}$ and $U$ with orthonormal columns the set $\mathcal{T}(S,U)$ (see Definition 12) is an appropriately defined set of $x \in \mathbb{R}^d$ that are ‘typical’ for $S$ and $U$. We first note that $q'$ is indeed the p.d.f. of a distribution. First, it is clear that $q'(x) \geq 0$ for all $x$. Second, we
have
\[
\int_{\mathbb{R}^d} q'(x)dx = \int_{\mathbb{R}^d} E_S \left[ \frac{e^{-\frac{1}{2} x^T M^{-1} x}}{\sqrt{(2\pi)^d \det M}} \cdot I[x \in T(S,U)] \right] + \xi \cdot \int_{\mathbb{R}^d} p(x)dx
\]
\[
= 1 - \int_{\mathbb{R}^d} E_S \left[ \frac{e^{-\frac{1}{2} x^T M^{-1} x}}{\sqrt{(2\pi)^d \det M}} \cdot I[x \notin T(S,U)] \right] + \xi
\]
\[
= 1 - \int_{\mathbb{R}^d \setminus T(S,U)} \frac{e^{-\frac{1}{2} x^T M^{-1} x}}{\sqrt{(2\pi)^d \det M}} dx + \xi
\]
\[
= 1 - E \left[ \Pr_{X \sim q}[X \notin T(S,U)] \right] + \xi
\]
\[
= 1, \quad \text{(by definition of } \xi)\]
as required.

As we show below, the total variation distance between $q'$ and $\tilde{q}$ is a small $n^{-10}$, so working with $q'$ suffices. The main argument of our proof shows that the distribution $q'(x)$ is close to $p(x)$ for 'typical' $x \in \mathbb{R}^d$. Then since $q'$ is close to $\tilde{q}$ and the event $E$ occurs with high probability, this suffices for a proof of Theorem 1. Formally, the core technical result behind the proof of Theorem 1 is

**Lemma 8.** There exists an absolute constant $C > 0$ such that for every $\delta \in (0,1)$ and every matrix $U \in \mathbb{R}^{n \times d}$ with orthonormal columns if $B \geq \frac{1}{4} C (\log n)^4 d^2$ there exists a set $E$ of CountSketch matrices and a subset $T^* \subseteq \mathbb{R}^d$ that satisfies $\Pr_{X \sim p}[X \notin T^*] \leq n^{-10}$ and $\Pr_{X \sim \tilde{q}}[X \notin T^*] \leq n^{-10}$ such that if $S \in \mathbb{R}^{B \times n}$ is a random CountSketch matrix, then (1) $\Pr_S[E] \geq 1 - \delta/3$, and (2) for all $x \in T^*$ one has

\[
\left| \frac{q'(x)}{p(x)} - 1 \right| \leq O((d^2 \log^4 n)/B) + O(n^{-10}).
\]

We now prove Theorem 1 assuming Lemma 8 and Claim 15. After this, we then prove Lemma 8 and Claim 15. We now give

**Proof of Theorem 1.** The proof relies on the observation that the vectors $(G_i SU)_{i=1}^m$ and $(\tilde{G}_i U)_{i=1}^m$ are vectors of independent samples from distributions $q(x)$ and $p(x)$ respectively. We denote the corresponding product distributions by $q^m$ and $p^m$. Since the good event $E$ constructed in Lemma 8 occurs with probability at least $1 - \delta/3$, it suffices to consider the distributions $\tilde{q}(x)$ and $p(x)$, as

\[
D_{TV}(q^m, p^m) \leq \Pr[E] + D_{TV}(\tilde{q}^m, p^m|E),
\]

where $D_{TV}(\tilde{q}^m, p^m|E)$ stands for the total variation distance between the distribution of $(\tilde{G}_i U)_{i=1}^m$ and the distribution of $(G_i SU)_{i=1}^m$ conditioned on $S \in E$. Further, we have by the triangle inequality

\[
D_{TV}(\tilde{q}^m, p^m|E) \leq D_{TV}((q^m)^m, p^m|E) + D_{TV}(\tilde{q}^m, (q^m)^m|E) \leq D_{TV}((q^m)^m, p^m|E) + m \cdot n^{-10},
\]

since $D_{TV}(\tilde{q}^m, (q^m)^m|E) \leq mD_{TV}(\tilde{q}, q^m|E) \leq mn^{-10}$, where $D_{TV}(\tilde{q}, q^m|E) \leq n^{-10}$ by Claim 15 below.

We first prove, using Lemma 8, that the KL divergence between $p(x)$ and $q'(x)$ restricted to the set $T^*$ (whose existence is guaranteed by Lemma 8) is bounded by $O((d \log n)^2/B^2)$. Specifically, let

\[
p_*(x) := \begin{cases} 
p(x)/\Pr_{X \sim p}[T^*] & \text{if } x \in T^* \\
0 & \text{o.w.} \end{cases}
\]

and

\[
q_*(x) := \begin{cases} 
q'(x)/\Pr_{X \sim q}[T^*] & \text{if } x \in T^* \\
0 & \text{o.w.} \end{cases}
\]
Since $T^*$ occurs with probability at least $1 - 1/n^{10}$ under both $\tilde{q}(x)$ and $p(x)$ by Lemma 19, it suffices to bound the total variation distance between the product of $m$ independent copies of $q^*_p(x)$ and $m$ independent copies of $p_*(x)$. Specifically,

$$D_{TV}((q^*)_m, p_m | \mathcal{E}) \leq D_{TV}((q^*)_m, p_m | (T^*)^m) + mPr[q^*_p(\mathbb{R}^d \setminus T^*)] + mPr[p(\mathbb{R}^d \setminus T^*)]$$

$$\leq D_{TV}((q^*)_m, p_m) + 2mn^{-10}, \quad \text{(by Lemma 19)}$$

(13)

where we used the fact that $q^*_p$ and $p_*$ are supported on $T^*$. Note that both distributions are still product distributions. By Pinsker’s inequality and the product structure we thus get

$$D_{TV}((q^*)_m, p_m) \leq \sqrt{\frac{1}{2} D_{KL}((q^*)_m || p_m^*)} \quad \text{(by Pinsker’s inequality)}$$

$$= \sqrt{\frac{m}{2} D_{KL}(q^*_p || p_*) \quad \text{(by additivity of KL divergence over product spaces)}}$$

(14)

In what follows we bound $D_{KL}(q^*_p || p_*)$. By Lemma 15 we have for every $x \in T^*$ that

$$|q^*_p(x)/p(x) - 1| \leq O((d^2 \log 4 n)/B) + O(n^{-10}),$$

(15)

so

$$|q^*_p(x)/p_*(x) - 1| = \left| \left(q^*_p(x)/p(x) \right) - \frac{\Pr_{X \sim q^*}[T^*]}{\Pr_{X \sim p}[T^*]} \right| - 1 = \left| \frac{\Pr_{X \sim q^*}[T^*]}{\Pr_{X \sim p}[T^*]} \right| \left| (q^*_p(x)/p(x)) - \frac{\Pr_{X \sim p}[T^*]}{\Pr_{X \sim q^*}[T^*]} \right|$$

$$\leq \frac{\Pr_{X \sim q}[T^*]}{\Pr_{X \sim p}[T^*]} \left( |q^*_p(x)/p(x) - 1| + \left| 1 - \frac{\Pr_{X \sim p}[T^*]}{\Pr_{X \sim q^*}[T^*]} \right| \right)$$

$$= (1 + O(n^{-10})) \cdot (|q^*_p(x)/p(x) - 1| + O(n^{-10}))$$

$$= O((d^2 \log 4 n)/B) + O(n^{-10}). \quad \text{(by (15)})$$

Since $B \geq \frac{1}{2} C d^2 \log^4 n$ for a sufficiently large constant $C > 0$ by assumption of the theorem, we get that

$$O((d^2 \log 4 n)/B) + O(n^{-10}) < O(1/C) + O(n^{-10}) < 1/2.$$

We thus get, using the bound $|1/(1 + x) - 1| \leq 2|x|$ for $|x| \leq 1/2$,

$$|p_*(x)/q^*_p(x) - 1| = \left| \frac{1}{q^*_p(x)/p_*(x)} - 1 \right| = \left| 1 + (q^*_p(x)/p_*(x) - 1) \right|$$

$$= O(|q^*_p(x)/p_*(x) - 1|)$$

$$= O((d^2 \log 4 n)/B) + O(n^{-10}) \quad \text{(16)}$$

We now use the fact that $|\ln(1 + x) - x| \leq 2x^2$ for all $x \in (-1/10, 1/10)$ to upper bound $D_{KL}(q^*_p || p_*)$. Specifically, we have

$$D_{KL}(q^*_p || p_*) = \mathbb{E}_{X \sim q^*_p} [\ln(q^*_p(X)/p_*(X))] \leq -\mathbb{E}_{X \sim q^*_p} [\ln(p_*(X)/q^*_p(X))]$$

$$\leq -\mathbb{E}_{X \sim q^*_p} [(p_*(x)/q^*_p(x)) - 1] -(p_*(x)/q^*_p(x) - 1)^2$$

$$\leq -\mathbb{E}_{X \sim q^*_p} [(p_*(x)/q^*_p(x) - 1)] + \mathbb{E}_{X \sim q^*_p} [(p_*(x)/q^*_p(x) - 1)^2]$$

$$= -(1 - 1) + \mathbb{E}_{X \sim q^*_p} [(p_*(x)/q^*_p(x) - 1)^2]$$

$$= \mathbb{E}_{X \sim q^*_p} [(p_*(x)/q^*_p(x) - 1)^2]$$

$$= O((d^2 \log 4 n)/B)^2 + n^{-10} \quad \text{(by (16))}$$

(17)

Since $B \geq \frac{1}{3} C (\log n)^4 d^2 \cdot m^{1/2}$ for a sufficiently large constant $C > 0$ by assumption of the theorem, substituting the bound of (17) into (14), we get

$$D_{TV}((q^*)_m, p_m^m) \leq \sqrt{\frac{m}{2}} \cdot D_{KL}(q^*_p || p_*) \leq \sqrt{\frac{m}{2}} \cdot O((d^2 \log 4 n)/B)^2 + n^{-10}) \leq \sqrt{\frac{m}{2}} \cdot \frac{\delta^2}{(8m)} \leq \frac{\delta}{2}.$$
Putting this together with [13], [10] and [9] using the assumption that \( m \leq n^4 \) gives the result. \( \square \)

The rest of the section is devoted to proving Lemma 8, i.e. bounding

\[
q'(x)/p(x) = E_S \left[ \exp \left( \frac{1}{2} x^T x - \frac{1}{2} x^T M^{-1} x - \frac{1}{2} \log \det M \right) \cdot I[x \in \mathcal{T}(S, U)] \right] + \xi, \tag{18}
\]

where \( \xi = E_S [ \Pr_{X \sim \mathcal{N}} [ X \notin \mathcal{T}(S, U)] | \mathcal{E} ] \leq n^{-20} \), for ‘typical’ \( x \) sampled from the Gaussian distribution (i.e. \( x \in \mathcal{T}^* \) – see formal definition below).

**Organization.** The rest of this section is organized as follows. We start by defining the set \( \mathcal{E} \) of ‘nice’ CountSketch matrices in section 1.1 and proving that a random CountSketch matrix is likely to be ‘nice’. We will in fact define a parameterized set \( \mathcal{E}(\gamma) \) in terms of a parameter \( \gamma \). In section 1.2 we define, for each matrix \( U \) (which can be thought of as fixed throughout our analysis) with orthonormal columns and CountSketch matrix \( S \), a set \( \mathcal{T}(S, U) \) of \( x \in \mathbb{R}^d \) that are ‘typical’ for \( S \) and \( U \). The ratio of pdfs in (18) can be approximated well by a Taylor expansion for such ‘typical’ \( x \in \mathcal{T}(S, U) \). These Taylor expansions are developed in section 1.3 and form the basis of our proof. Unfortunately, these Taylor expansions are valid only for \( x \in \mathcal{T}(S, U) \), i.e. for \( x \) that are ‘typical’ with respect to a given \( S \). To complete the proof, we need to construct a universal ‘typical’ set \( \mathcal{T}^*(U, \gamma) \) of \( x \in \mathbb{R}^d \), again parameterized in terms of a parameter \( \gamma \), that will allow for approximation via Taylor expansions for all \( x \in \mathcal{T}^*(U, \gamma) \) and \( S \in \mathcal{E}(\gamma) \). We construct such a set \( \mathcal{T}^*(U, \gamma) \) in section 1.4. Finally, the proof of Lemma 8 is given in section 1.5.

### 1.1 Typical set \( \mathcal{E} \) of CountSketch matrices and its properties

Our analysis of (18) starts by Taylor expanding \( M^{-1} \) and \( \log \det M \) around the identity matrix. We now state the Taylor expansions, and the define a (family of) high probability events \( \mathcal{E}(\gamma) \) (equivalently, sets of ‘typical’ CountSketch matrices) such that the Taylor expansions are valid for matrices \( M \in \mathcal{E}(\gamma) \) for all sufficiently small \( \gamma \)

1. The Taylor expansions that we use are given by

\[
E_S [ \Pr_{X \sim \mathcal{N}} [ X \notin \mathcal{T}(S, U)] | \mathcal{E} ] \leq n^{-20},
\]

for ‘typical’ \( x \) sampled from the Gaussian distribution (i.e. \( x \in \mathcal{T}^* \) – see formal definition below).

- **Claim 9.** For any matrix \( M \) with \( ||I - M|| < 1/2 \) one has
  \[
  M^{-1} = (I - (I - M))^{-1} = \sum_{k \geq 0} (I - M)^k.
  \]

- **Claim 10.** For any matrix \( M \) with \( ||I - M|| < 1/2 \) one has
  \[
  \log \det M = \log \det (I - (I - M)) = \sum_{k \geq 1} -\text{Tr}((I - M)^k)/k.
  \]

For a parameter \( \gamma \in (0, 1) \) that we will later set to 1/poly(log \( n \)), define event \( \mathcal{E}(\gamma) \) as

\[
\mathcal{E}(\gamma) := \{ ||I - M||_F^2 \leq \gamma^2 \quad \text{and} \quad |\text{Tr}(I - M)| \leq \gamma \}.
\]

The events \( \mathcal{E}(\gamma) \) occur with high probability even for fairly small \( \gamma \) as long as \( B \) is sufficiently large:

- **Claim 11.** For any matrix \( U \in \mathbb{R}^{n \times d} \) with orthonormal columns, any \( B \times n \) CountSketch matrix \( S \) we have
  \[
  \Pr[\mathcal{E}(\gamma)] \geq 1 - 3(d/\gamma)^2/B.
  \]

**Proof.** By Lemma 21 below, we have

\[
E_S[||I - M||_F^2] \leq 2d^2/B.
\]

Applying Markov’s inequality to \( ||I - M||_F^2 \), we get

\[
\Pr[||I - M||_F^2 \geq \gamma^2] \leq \Pr[||I - M||_F^2 \geq \gamma^2(B/(2d^2))] \leq 2(d/\gamma)^2/B
\]

as required.

We also have by Lemma 21 (fifth bound) that

\[
E_S[(\text{Tr}(I - M))^2] \leq d^2/B.
\]

Applying Markov’s inequality to \((\text{Tr}(I - M))^2\), we get

\[
\Pr[|\text{Tr}(I - M)| \geq \gamma] = \Pr[(\text{Tr}(I - M))^2 \geq \gamma^2] \leq \Pr[(\text{Tr}(I - M))^2 \geq \gamma^2(B/(d^2))] \leq (d/\gamma)^2/B.
\]

A union bound over the two events gives the result. \( \square \)

---

Note that we use the notation \( S \in \mathcal{E}(\gamma) \) and \( M \in \mathcal{E}(\gamma) \) interchangeably. This is fine since \( M = U^T S^T S U \) and the matrix \( U \) is fixed.
1.2 Typical sets $T(S,U)$ and their properties

In order to construct a single typical set $T^*$, we will need the following simple definitions of sets $T(S,U)$ of $x \in \mathbb{R}^d$ that are ‘typical’ for a given CountSketch matrix (as opposed to the set $T^*$ whose existence is guaranteed by Lemma 8, which contains $x$ that are ‘typical’ for all matrices $S \in E$ simultaneously). We will use

**Definition 12** (Typical $x$). For any orthonormal matrix $U \in \mathbb{R}^{n \times d}$ and CountSketch matrix $S$ we define

$$T(S,U) := \left\{ x \in \mathbb{R}^d : |x^T(I - M)x| \leq \frac{1}{100} \text{ and } |x^T(I - M)^2x| \leq \frac{1}{100} \right\}.$$ 

The following claim will be useful in what follows. Its (simple) proof is given in the appendix:

**Claim 13.** For any matrix $U \in \mathbb{R}^{n \times d}$ with orthonormal columns and any CountSketch matrix $S \in \mathbb{R}^{B \times n}$ one has $||I - M||_F^2 \leq 4n^3$.

The following claim is crucial to our analysis. A detailed proof is given in the appendix.

**Claim 14.** For any matrix $U \in \mathbb{R}^{n \times d}$ with orthonormal columns, every $\gamma \leq 1/\log^2 n$, every CountSketch matrix $S \in E(\gamma)$ one has (1) $Pr_{X \sim N(0, I_d)}[X \notin T(S,U)] < n^{-40}$ and (2) for any CountSketch matrix $S' \in E(\gamma)$, $M' = U^T S'^T S'U$ one has $Pr_{X \sim N(0, M')}[X \notin T(S,U)] < n^{-40}$ for sufficiently large $n$.

Using the claim above we get

**Claim 15.** The total variation distance between $\tilde{q}$ (defined in (7)) and $q'$ (defined in (8)) is at most $n^{-10}$. Further, $\xi \leq n^{-10}$.

**Proof.** We have

$$D_{TV}(\tilde{q}, q') \leq 2\xi \leq 2 \int_{\mathbb{R}^d} E_S \left[ \frac{1}{\sqrt{(2\pi)^d \det M}} e^{-\frac{1}{2} x^T M^{-1}x} \cdot I[x \notin T(S,U)] \right] dx$$

$$= 2E_S \left[ \int_{\mathbb{R}^d} \frac{1}{\sqrt{(2\pi)^d \det M}} e^{-\frac{1}{2} x^T M^{-1}x} \cdot I[x \notin T(S,U)] dx \right] \mathbb{E}(\gamma)$$

$$= 2E_S \left[ Pr_{X \sim N(0, M)}[X \notin T(S,U)] \right] \mathbb{E}(\gamma)$$

$$\leq 2n^{-40} \leq n^{-10} \quad \text{(by Claim 14)}$$

as required.

1.3 Basic Taylor expansions

In this section we define the basic Taylor expansions of $\bar{q}(x)/p(x)$ that form the foundation of our analysis. Our analysis of (18) proceeds by first Taylor expanding $M^{-1}$ and $\det M$ around the identity matrix using Claims 9 and 10 which is valid since for any $S \in E(\gamma)$ for $\gamma < 1/2$ one has $||I - M||_2 \leq ||I - M||_F \leq 1/2$. This gives

$$\bar{q}(x)/p(x) = E_S \left[ \exp \left( -\frac{1}{2} x^T(I - M)x + \frac{1}{2} \sum_{k \geq 2} (x^T(I - M)^k x) / k \right) \right]$$

$$= E_S \left[ \exp \left( -\frac{1}{2} x^T(I - M)x + \frac{1}{2} \sum_{k \geq 1} (x^T(I - M)^k x - \Tr((I - M)^k)/k) \right) \right]$$

$$= E_S \left[ \exp \left( -\frac{1}{2} x^T(I - M)x + \frac{1}{2} \Tr(I - M) - R(x) \right) \right],$$

(20)
where \( R(x) := \frac{1}{2} \sum_{k \geq 2} (x^T(I - M)^k x - \text{Tr}((I - M)^k)/k). \)

The rationale behind the definition of \( \mathcal{E}(\gamma) \) is that for all \( S \in \mathcal{E}(\gamma) \) the residual \( R(x) \) above is (essentially) dominated by the quadratic terms, i.e. \( ||I - M||^2_2 \) and \( x^T(I - M)^2 x \) (for ‘typical’ values of \( x \) – see Lemma 18 below), i.e. we can truncate the Taylor expansion to the first and second terms and control the error. This is made formal by the following three lemmas.

**Lemma 16.** For every \( \gamma \in (0, 1) \), conditioned on \( \mathcal{E}(\gamma) \) we have \( \text{Tr}((I - M)^k) \leq \gamma^{k-2} \cdot ||I - M||^2_F \) for all \( k \geq 2 \).

**Proof.** \(|\text{Tr}((I - M)^k)| \leq ||I - M||^2_{2} \cdot \text{Tr}((I - M)^2) \leq ||I - M||^2_{2} \cdot ||I - M||^2_F \), as required, since \( ||A||_2 \leq ||A||_F \) and \( \text{Tr}(A^T A) = ||A||^2_2 \) for all \( A \in \mathbb{R}^{d \times d} \).

**Lemma 17.** For any matrix \( U \in \mathbb{R}^{n \times d} \) with orthonormal columns, any \( \gamma \in (0, 1/2) \), for any \( x \in \mathbb{R}^d \) one has, for any CountSketch matrix \( S \in \mathcal{E}(\gamma) \), \( x^T(I - M)^k x \leq \gamma^{k-2} x^T(I - M)^2 x \) for any \( k \geq 2 \).

**Proof.** We have, for any \( x \in \mathbb{R}^d \) and any \( S \in \mathcal{E}(\gamma) \), \( |x^T(I - M)^k x| \leq ||I - M||^2_{2} \cdot x^T(I - M)^2 x \leq \gamma^{k-2} \cdot x^T(I - M)^2 x \), as \( ||I - M||^2_2 \leq ||I - M||_F \).

**Lemma 18.** For any \( \gamma \in (0, 1/2) \), any matrix \( U \in \mathbb{R}^{n \times d} \) with orthonormal columns, any CountSketch matrix \( S \in \mathcal{E}(\gamma) \) and any \( x \in \mathcal{T}(S, U) \) one has

\[
|R(x)| = \sum_{k \geq 2} |x^T(I - M)^k x| + |\text{Tr}((I - M)^k)|/k \leq C||I - M||^2_F + Cx^T(I - M)^2 x,
\]

where \( C > 0 \) is an absolute constant.

**Proof.** We have by combining Lemma 16 and Lemma 17

\[
\sum_{k \geq 2} |x^T(I - M)^k x| + |\text{Tr}((I - M)^k)|/k \leq \sum_{k \geq 2} \gamma^{k-2} x^T(I - M)^2 x + \gamma^{k-2} \cdot ||I - M||^2_F/k
\]

\[
\leq C(x^T(I - M)^2 x + ||I - M||^2_F)
\]

for an absolute constant \( C' > 0 \), as \( \gamma < 1/2 \) by assumption of the lemma.

**1.4 Constructing the universal set \( \mathcal{T}^*(U, \gamma) \) of typical \( x \)**

The main result of this section is the following lemma:

**Lemma 19.** For every matrix \( U \in \mathbb{R}^{n \times d} \) with orthonormal columns, for every \( \gamma \in (0, 1/\log^2 n) \) and any \( \delta > 0 \) if

\[
\mathcal{T}^*(U, \gamma) := \{ x \in \mathbb{R}^d \text{ s.t. } ||x||_\infty \leq C\sqrt{\log n} \text{ and } ||(Ux)_a|| \leq O(\sqrt{\log n})||U_a||_2 \text{ for all } a \in [n] \text{ and } E_S[I[x \notin \mathcal{T}(S, U)]|\mathcal{E}(\gamma)] < 1/n^{25}.\}
\]

then (a) \( \Pr_{X \sim N(0, I_d)}[X \in \mathcal{T}^*(U, \gamma)] \geq 1 - n^{-10} \) and (b) \( \Pr_{X \sim \tilde{q}}[X \in \mathcal{T}^*(U, \gamma)] \geq 1 - n^{-10}. \)

Note that the lemma guarantees the existence of a universal set \( \mathcal{T}^* \subseteq \mathbb{R}^d \) that captures most of the probability mass of both the normal distribution \( N(0, I_d) \) and the mixture \( \tilde{q} \).

**Proof of Lemma 19**

Let \( \mathcal{T}_1^* := \{ x \in \mathbb{R}^d : E_S[I[x \notin \mathcal{T}(S, U)]|\mathcal{E}(\gamma)] < 1/n^{25}.\} \), \( \mathcal{T}_2^* := \{ x \in \mathbb{R}^d : ||x||_\infty \leq C\sqrt{\log n} \}, \) and \( \mathcal{T}_3^* := \{ x \in \mathbb{R}^d : ||(Ux)_a|| \leq C\sqrt{\log n}||U_a||_2 \text{ for all } a \in [n] \}. \)

We prove that \( \mathcal{T}_i^*, i = 1, 2, 3 \) occur with high probability under both distributions. As we show below, the result then follows by a union bound.
Showing that $T_1^*$ occurs with high probability. We first show that $T_1^*$ occurs with high probability under the isotropic Gaussian distribution $X \sim N(0, I_d)$, and then show that it also occurs with high probability under the mixture of Gaussians distribution $\tilde{q}$. In both cases the proof proceeds by applying Claim 14 followed by Markov’s inequality.

Step 1: bounding $\Pr_{X \sim N(0,I_d)}[T_1^*]$. We have by Claim 14 (1) that $\Pr_{X \sim N(0,I_d)}[I[X \not\in T(S,U)]] < n^{-40}$, and hence

$$E_S \left[ E_{X \sim N(0,I_d)}[I[X \not\in T(S,U)]] \right] < 1/n^{40},$$

implying that $E_{X \sim N(0,I_d)}[E_S[I[X \not\in T(S,U)]]|\mathcal{E}(\gamma)] < 1/n^{40}$. We thus get by Markov’s inequality that $\Pr_{X \sim N(0,I_d)}[T_1^*] \geq 1 - n^{-15}$.

Step 2: bounding $\Pr_{X \sim \tilde{q}}[T_1^*]$. We have by Claim 14 (2) that for any $U \in \mathbb{R}^{n \times d}$ with orthonormal columns, any pair of matrices $S, S' \in \mathcal{E}(\gamma)$, if $M' = UT^TSU$, then $\Pr_{X \sim N(0,M')}[X \not\in T(S,U)] < n^{-40}$. We thus have

$$E_{X \sim \tilde{q}}[E_S[I[X \not\in T(S,U)]]|\mathcal{E}(\gamma)] = E_{S'}[E_{X \sim qS'}[E_S[I[X \not\in T(S,U)]]|\mathcal{E}(\gamma)]] \leq n^{-40}.$$

By Markov’s inequality applied to the expression in the first line we thus have

$$\Pr_{X \sim \tilde{q}}[E_S[I[X \not\in T(S,U)]]|\mathcal{E}(\gamma)] < 1/n^{25} < n^{-15}.$$

Showing that $T_2^*$ occurs with high probability. The fact that

$$\Pr_{X \sim N(0,I_d)}[\|X\|_\infty \leq C\sqrt{\log n}] \geq 1 - n^{-40}$$

follows by standard properties of Gaussian random variables. Thus, it remains to show that $T_2^*$ occurs with high probability under $X \sim \tilde{q}$. For any $U \in \mathbb{R}^{n \times d}$ and $S \in \mathcal{E}(\gamma)$ we now prove that for $M = UT^TSU$

$$\Pr_{X \sim N(0,M)}[\|X\|_\infty \leq C\sqrt{\log n}] \geq 1 - n^{-40} \quad (21)$$

It is convenient to let $X = M^{1/2}Y$, where $Y \sim N(0,I_d)$ is a vector of independent Gaussians of unit variance. Then we need to bound

$$\Pr_{X \sim N(0,M)}[\|X\|_\infty \geq C\sqrt{\log n}] = \Pr_{Y \sim N(0,I_d)}[\|M^{1/2}Y\|_\infty \geq C\sqrt{\log n}]$$
Lemma 20. There exists an absolute constant $C > 0$ such that for every $\gamma \in (0, 1/\log n)$, any matrix $U \in \mathbb{R}^{n \times d}$ with orthonormal columns and any CountSketch matrix $S \in \mathcal{E}(\gamma)$ and $x \in \mathcal{T}(S,U)$ one has, letting

$$L(x) := -\frac{1}{2}x^T(I - M)x + \frac{1}{2}\text{Tr}(I - M) - \frac{1}{8}x^T(I - M)x \cdot \text{Tr}(I - M)$$

By 2-stability of the Gaussian distribution we have that for each $i = 1, \ldots, d$ the random variable $(M^{1/2}Y)_i$ is Gaussian with variance at most $||M^{1/2}||_F^2$, which we bound by

$$||M^{1/2}||_F = ||(I + (M - I))^{1/2}||_F = \left\| \sum_{t=0}^{\infty} \binom{1/2}{t} (I - M)^t \right\|_F$$

$$\leq \sum_{t=0}^{\infty} \left( \frac{1/2}{t} \right) \cdot ||(I - M)^t||_F$$

$$\leq \sum_{t=0}^{\infty} \left( \frac{1/2}{t} \right) \cdot ||I - M||_F^t$$

$$\leq \sum_{t=0}^{\infty} ||I - M||_F^t$$

$$\leq 2$$

Thus, for each $i \in [n]$ the random variable $(M^{1/2}Y)_i$ is Gaussian with variance at most 4, and follows by standard properties of Gaussian random variables as long as $C > 0$ is a sufficiently large constant.

**Showing that $T^*_3$ occurs with high probability.** The fact that

$$\Pr_{X \sim N(0, I_d)} \left[ |(UX)_a| \leq C\sqrt{\log n} \cdot ||U_a||_2 \right. \text{ for all } a \in [n] \left. \right] \geq 1 - n^{-40}$$

follows by standard properties of Gaussian random variables and a union bound over all $a \in [n]$.

Thus, it remains to show that $T^*_3$ occurs with high probability under $X \sim \tilde{q}$. For any $U \in \mathbb{R}^{n \times d}$ and $S \in \mathcal{E}(\gamma)$ we now prove that for $M = U^T S^T SU$

$$\Pr_{X \sim N(0, M)} \left[ |(UX)_a| \leq C\sqrt{\log n} ||U_a||_2 \right. \text{ for all } a \in [n] \left. \right] \geq 1 - n^{-40}$$

It is convenient to let $X = M^{1/2}Y$, where $Y \sim N(0, I_d)$ is a vector of independent Gaussians of unit variance. Then we need to bound, for each $a \in [n]$

$$\Pr_{X \sim N(0, M)} \left[ |(UX)_a| \geq C\sqrt{\log n} ||U_a||_2 \right) = \Pr_{Y \sim N(0, I_d)} \left[ |(UM^{1/2}Y)_a| \geq C\sqrt{\log n} ||U_a||_2 \right]$$

By 2-stability of the Gaussian distribution we have that for each $a = 1, \ldots, n$ the random variable $U_aM^{1/2}Y$ is Gaussian with variance at most $||U_aM^{1/2}||^2_2 \leq 4||U_a||^2_2$ (since $\gamma < 1/\log^2 n$ by assumption of the lemma), and hence the result follows by standard properties of Gaussian random variables and a union bound.

Finally, we let $T^* := T^*_1 \cap T^*_2 \cap T^*_3$. By a union bound applied to the bounds above we have that $T^*$ occurs with probability at least $1 - n^{-10}$ under both distributions, as required.

1.5 Proof of Lemma 8

We first prove

**Lemma 20.** There exists an absolute constant $C > 0$ such that for every $\gamma \in (0, 1/\log n)$, any matrix $U \in \mathbb{R}^{n \times d}$ with orthonormal columns and any CountSketch matrix $S \in \mathcal{E}(\gamma)$ and $x \in T(S,U)$ one has,
and
\[ Q(x) := ((x^T (I - M)x)^2 + (\text{Tr}(I - M))^2 + x^T (I - M)^2 x + \|I - M\|_F^2), \]
that
\[ |1 + L(x) - \exp \left( \frac{1}{2} x^T x - \frac{1}{2} x^T M^{-1} x - \frac{1}{2} \log \det M \right) \| \leq C \cdot Q(x). \]

The proof is given in section \textsection A.

We will need the following two lemmas, whose proofs are provided in section A.2.

**Lemma 21.** For any \( U \in \mathbb{R}^{n \times d} \) with orthonormal columns, and \( B \geq 1 \), if \( S \) is a random CountSketch matrix and \( M = U^T S^T S U \), then

1. \( E_S \|M - I\|_F^2 \leq 2d^2 / B \)
2. for all \( x \in T^* \) one has \( E_S [x^T (I - M)^2 x] = O(d^2 (\log^2 n) / B) \)
3. for all \( x \in T^* \) one has \( E_S [(x^T (I - M) x)^2] = O(d^2 (\log^2 n) / B) \)
4. for all \( x \in T^* \) one has \( E_S [(x^T (I - M) x) \cdot \text{Tr}(I - M)] = O(d^2 (\log n) / B) \)
5. one has \( E_S [(\text{Tr}(I - M))^2] = O(d^2 / B) \)

and

**Lemma 22** (Variance bound). For any matrix \( U \in \mathbb{R}^{n \times d} \) with orthonormal columns if \( \gamma \in (0, 1/2) \) and \( T^*(U, \gamma) \subseteq \mathbb{R}^d \) as defined in Lemma \textsection 19, then for any \( x \in T^*(U, \gamma) \) one has, for

\[ L(x) := -\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) - \frac{1}{8} x^T (I - M)x \cdot \text{Tr}(I - M) \]

and
\[ Q(x) := ((x^T (I - M)x)^2 + (\text{Tr}(I - M))^2 + x^T (I - M)^2 x + \|I - M\|_F^2), \]
that for any constant \( C \)
\[ E_S \left[ (L(x) + C \cdot Q(x))^2 \right] = O(d^2 (\log^2 n) / B), \]
where \( S \) is a uniformly random CountSketch matrix and \( M = U^T S^T S U \).

We will use the following lemma, whose proof is given in section \textsection A.

**Lemma 23.** For any random variable \( Z \) and any event \( \mathcal{E} \) with \( \Pr[\mathcal{E}] \geq 1/2 \), if \( \epsilon := E[(Z - 1)^2] \), then
\[ |E[Z] - E[Z|\mathcal{E}]| \leq 2(1 + E[Z|\mathcal{E}])\Pr[\mathcal{E}] + 2\sqrt{\epsilon \Pr[\mathcal{E}]} . \]

Equipped with the bounds above, we can now prove Lemma \textsection 8.

**Lemma 8 (Restated)** There exists an absolute constant \( C > 0 \) such that for every \( \delta \in (0, 1) \) and every matrix \( U \in \mathbb{R}^{n \times d} \) with orthonormal columns if \( B \geq \frac{1}{2} C (\log n)^3 d^2 \) there exists a set \( \mathcal{E} \) of CountSketch matrices and a subset \( T^* \subseteq \mathbb{R}^d \) that satisfies \( \Pr_{X \sim p}[X \notin T^*] \leq n^{-10} \) and \( \Pr_{X \sim q}[X \notin T^*] \leq n^{-10} \) such that if \( S \in \mathbb{R}^{B \times n} \) is a random CountSketch matrix, then (1) \( \Pr_S[\mathcal{E}] \geq 1 - \delta / 3 \), and (2) for all \( x \in T^* \) one has
\[ \left| \frac{q'(x)}{p(x)} - 1 \right| \leq O((d^2 \log^4 n) / B) + O(n^{-10}) . \]
Proof. Let $\mathcal{T}^*(U, \gamma) \subseteq \mathbb{R}^d$ be as defined in Lemma 19, and let $\gamma := 1/\log^2 n$. Let $E := \mathcal{E}(\gamma)$, and note that $\Pr[E] \geq 1 - \delta/3$ by Claim 11 as long as $C$ is a large enough constant, as required.

We now bound
\[
q'(x) = \mathbb{E}_S \left[ \frac{qS(x)}{p(x)} \cdot I[x \in \mathcal{T}(S, U)] \left| \mathcal{E}(\gamma) \right. \right] + \xi,
\]
for $x \in \mathcal{T}^*(U, \gamma)$, where $\xi = \mathbb{E}_S[\Pr_{x \sim qS}[X \in \mathcal{T}(S, U)]] \leq n^{-40}$ by definition and Claim 13 (2). For each $S \in \mathcal{E}(\gamma)$ and $x \in \mathcal{T}(S, U)$ we have by Lemma 20
\[
| \frac{qS(x)}{p(x)} - (1 + L(x)) | = \left| \exp \left( \frac{1}{2} x^T x - \frac{1}{2} x^T M^{-1} x - \frac{1}{2} \log \det M \right) - (1 + L(x)) \right| \leq C \cdot Q(x),
\]
where
\[
L(x) := -\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) - \frac{1}{8} x^T (I - M)x \cdot \text{Tr}(I - M)
\]
denotes the ‘linear’ term and
\[
Q(x) := (x^T (I - M)x)^2 + (\text{Tr}(I - M))^2 + x^T (I - M)^2 x + ||I - M||_F^2
\]
denotes the ‘quadratic’ term.

Taking expectations, we get
\[
\mathbb{E}_S \left[ (L(x) - C \cdot Q(x)) \cdot I[x \in \mathcal{T}(S, U)] \left| \mathcal{E}(\gamma) \right. \right] \leq \mathbb{E}_S \left[ \left( \exp \left( \frac{1}{2} x^T x - \frac{1}{2} x^T M^{-1} x - \frac{1}{2} \log \det M \right) - 1 \right) \cdot I[x \in \mathcal{T}(S, U)] \left| \mathcal{E}(\gamma) \right. \right]
\]
\[
\mathbb{E}_S \left[ (L(x) + C \cdot Q(x)) \cdot I[x \in \mathcal{T}(S, U)] \left| \mathcal{E}(\gamma) \right. \right] = O((Cd\log n)^2/B) + O(n^{-10}),
\]
which we do now. We only provide the analysis for the case when the sign in front of the constant $C$ is a plus, as the other part is analogous.

We first show that removing the multiplier $I[x \in \mathcal{T}(S, U)]$ from the equation above only changes the expectation slightly. Specifically, note that
\[
\mathbb{E}_S \left[ (L(x) + C \cdot Q(x)) \cdot I[x \in \mathcal{T}(S, U)] \left| \mathcal{E}(\gamma) \right. \right] - \mathbb{E}_S \left[ L(x) + C \cdot Q(x) \left| \mathcal{E}(\gamma) \right. \right]
\]
\[
\leq \mathbb{E}_S \left[ |L(x) + C \cdot Q(x)| \cdot I[x \notin \mathcal{T}(S, U)] \left| \mathcal{E}(\gamma) \right. \right].
\]

By Claim 13 we have $||I - M||_F^2 \leq 4n^3$ for all $S$ and $U$, so every element of the matrix $I - M$ is upper bounded by $2n^2$. Similarly, we have $||(I - M)^2||_F \leq ||I - M||_F^2$, and so every element of $(I - M)^2$ is upper bounded by $4n^3$. Thus, for any $x \in \mathcal{T}^*(U, \gamma)$ one has
\[
|L(x) + CQ(x)|
\]
\[
\leq (|x^T (I - M)x| + |\text{Tr}(I - M)| + |x^T (I - M)x \cdot \text{Tr}(I - M)|
\]
\[
+ C((x^T (I - M)x)^2 + (\text{Tr}(I - M))^2 + x^T (I - M)^2 x + ||I - M||_F^2))
\]
\[
= O(\log n)(2n^2d^2 + d \cdot (2n^2) + (2n^2)d^2 + (2n^2)^2d^2 + (d \cdot 2n^2)^2 + 4n^4d^2 + 4n^3) \leq n^{10}
\]
as long as $n$ is sufficiently large, where we used the fact that $||x||_{\infty} \leq O(\sqrt{\log n})$ for all $x \in \mathcal{T}^*(U, \gamma)$.

Furthermore, by Lemma 19 we have for $x \in \mathcal{T}^*(U, \gamma)$ that
\[
\mathbb{E}_S \left[ I[x \notin \mathcal{T}(S, U)] \left| \mathcal{E}(\gamma) \right. \right] < 1/n^{25}.
\]
Substituting these two bounds into (22), we get

$$E_S[|L(x) + C \cdot Q(x)| \cdot I[x \notin T(S,U)] E(\gamma)] \leq n^{-10}$$

(23)

so it remains to bound

$$E_S[L(x) + C \cdot Q(x)] E(\gamma).$$

We bound the expectation above by relating it to the corresponding unconditional expectation. Let

$$Z := 1 + (L(x) + C \cdot Q(x)),$$

and note that

$$E[Z] = 1 - E_S\left[\frac{1}{S} x^T (I - M)x \cdot Tr(I - M)\right] + C \cdot E_S[Q(x)] = 1 + O((C \log n)^2 d^2/B)$$

by Lemma 21. Let $$\epsilon := E_S([Z - 1]^2]$$. We note that by Lemma 22 that $$\epsilon \leq O((d^2 (\log n)^2)/B),$$ and hence since $$E(\gamma) \geq 1/2$$ by Claim 11 by Lemma 23 we have

$$|E[Z] - E[Z | E(\gamma)|] \leq 2(1 + E[Z]) \Pr[E(\gamma)] + 2\sqrt{c \Pr[E(\gamma)]}.$$}

Since $$\Pr[E(\gamma)] \leq 3(d/\gamma)^2/B$$ by Claim 11 and using the assumption that $$B \geq (\log n)d^2$$, we get

$$|E[Z] - E[Z | E(\gamma)|] \leq O((d/\gamma)^2/B) + 2\sqrt{O(d^2 \log n/B) \cdot (d/\gamma)^2/B} = O((\frac{1}{\gamma^2} + \frac{1}{\gamma} \log n)d^2/B) = O((d/\gamma)^2/B),$$

(25)

where we used the assumption that $$\gamma \leq 1/\log^2 n$$. Combining (25), (24) with (22) and (23), we get

$$\left| \frac{g'(x)}{p(x)} - 1 \right| = \left| E_S \left[ \frac{g(x)}{p(x)} \cdot I[x \in T(U,S)] E(\gamma) \right] + \xi - 1 \right| \leq O((d^2 \log^4 n)/B) + O(1/n^{10}).$$

A Proofs omitted from the main body

A.1 Proof of Claim 14 and Claim 13

We will use

Theorem 24 (Bernstein’s inequality). Let $$X_1, \ldots, X_n$$ be independent zero mean random variables such that $$|X_i| \leq L$$ for all $$i$$ with probability 1, and let $$X := \sum_{i=1}^n X_i$$. Then

$$\Pr[X > t] \leq \exp \left( -\frac{\frac{1}{2} t^2}{\sum_{i=1}^n E[|X_i|^2] + \frac{1}{2} Lt} \right).$$

Proof of Claim 14

Proving (1). The bound follows by standard concentration inequalities, as we now show. Since the normal distribution is rotationally invariant, we have that

$$X^T (I - M) X = \sum_{i=1}^d (\lambda_i - 1) Y_i^2 = Tr(M - I) + \sum_{i=1}^d (\lambda_i - 1)(Y_i^2 - 1),$$

(26)

where $$Y \sim N(0, I_d)$$ and $$\lambda_i$$ are the eigenvalues of $$M$$. We now apply Bernstein’s inequality (Theorem 24) to random variables $$(\lambda_i - 1)(Y_i^2 - 1)$$ (note that they are zero mean). We also have $$E[(\lambda_i - 1)^2(Y_i^2 - 1)^2] \leq$$
We have that $\sum_{i=1}^{d} (\lambda_i - 1)(Y_i^2 - 1)/2 \cdot \frac{1}{100}$, for all $S \subseteq E(\gamma)$ to obtain the result. We also have $|I - M|_{\ell^2} C \sqrt{\log n} \leq \gamma \cdot C \sqrt{\log n}$ for all $i$ with probability at least $1 - n^{-40}/4$ as long as $C > 0$ is larger than an absolute constant. We thus have by applying Theorem 14 to random variables clipped at $\gamma C \sqrt{\log n}$ in magnitude, which we denote by event $\mathcal{F}$, to conclude for all $t \geq 0$,

$$\Pr\left[\sum_{i=1}^{d} (\lambda_i - 1)(Y_i^2 - 1)/2 > t \mid \mathcal{F}\right] < 2 \exp\left(-\frac{\frac{1}{2}t^2}{O(\sum_{i=1}^{n}(\lambda_i - 1)^2) + (\frac{1}{2}\gamma C \sqrt{\log n})t}\right).$$

Note the random variables are still independent and zero-mean conditioned on $\mathcal{F}$, and $\mathbb{E}[(\lambda_i - 1)^2(Y_i^2 - 1)] \leq O((\lambda_i - 1)^2)$ continues to hold, since the clipping changes the expectation by at most a factor of $(1 + O(n^{-40}))$. By a union bound we can remove the conditioning on $\mathcal{F}$,

$$\Pr\left[\sum_{i=1}^{d} (\lambda_i - 1)(Y_i^2 - 1)/2 > t\right] < 2 \exp\left(-\frac{\frac{1}{2}t^2}{O(\sum_{i=1}^{n}(\lambda_i - 1)^2) + (\frac{1}{2}\gamma C \sqrt{\log n})t}\right) + \frac{n^{-40}}{4}.$$

Setting $t = \frac{1}{100}$, and using the fact that $\sum_{i}(\lambda_i - 1)^2 = ||I - M||_{\ell^2}^2 \leq \gamma^2$, we get

$$\Pr\left[\sum_{i=1}^{d} (\lambda_i - 1)(Y_i^2 - 1)/2 > \frac{1}{2} \cdot \frac{1}{100}\right] < 2 \exp\left(-\frac{\frac{1}{2}(\frac{1}{2} \cdot \frac{1}{100})^2}{O(\gamma^2) + (\frac{1}{2} \cdot \frac{1}{100})\gamma C \sqrt{\log n}}\right) + \frac{n^{-40}}{4} = \exp(-\Omega(1/\gamma^2 \log n)) + \frac{n^{-40}}{4} < \frac{n^{-40}}{2},$$

since $\gamma \leq 1/\log^2 n$ by assumption, for a sufficiently large $n$. Combining this with (26), we get, using the fact that $|\text{Tr}(I - M)| \leq \gamma < \frac{1}{2} \cdot \frac{1}{100}$ for $S \subseteq E(\gamma)$ that

$$\Pr[X^T(I - M)X > \frac{1}{100}] \leq \Pr[\sum_{i=1}^{d} (\lambda_i - 1)(Y_i^2 - 1)/2 > \frac{1}{2} \cdot \frac{1}{100}] < \frac{n^{-40}}{2},$$

as required.

We also have

$$X^T(I - M)^2X = \sum_{i=1}^{d} (\lambda_i - 1)^2Y_i^2 \leq ||I - M||_{\ell^2}^2 \cdot \max_{i \in [d]} |Y_i|^2 \leq O(\log n) \cdot ||I - M||_{\ell^2}^2 = O(\log n \gamma^2) \leq \frac{1}{100}$$

with probability at least $1 - n^{-40}/2$ by standard properties of Gaussian random variables. Putting the two estimates together and taking a union bound over the failure events now shows that $\Pr_{X \sim N(0, I_d)}[X \notin T(S, U)] < n^{-40}$, as required.

**Proving (2).** Recall that $T(S, U) = \{x \in \mathbb{R}^d : |x^T(I - M)x| \leq \frac{1}{100}$ and $x^T(I - M)^2x \leq \frac{1}{100}\}$. For any $S'$ we have that $X \sim N(0, M')$, where $M' = (S'U)^T S'U$, so $X = M'^{1/2}Y$, where $Y = N(0, I_d)$. We thus have

$$X^T(I - M)X = (M'^{1/2}Y)^T(I - M)(M'^{1/2}Y) = Y^T M'^{1/2}(I - M)M'^{1/2}Y.$$

We now show that

$$\Pr_{Y \sim N(0, I_d)}\left[|Y^T M'^{1/2}(I - M)M'^{1/2}Y| > \frac{1}{100}\right] < 1/n^{20}$$

(27)
Let \( Q := M'^{1/2}(I - M)M'^{1/2} \), and let \( 1 - \tilde{\lambda}_i, i = 1, \ldots, d \) denote the eigenvalues of \( Q \). We have

\[
Y^T M'^{1/2}(I - M)M'^{1/2}Y = \sum_{i=1}^{d} (1 - \tilde{\lambda}_i)Z_i^2,
\]

where \( Z \sim N(0, I_d) \). Note that

\[
\left| \sum_{i=1}^{d} (1 - \tilde{\lambda}_i) \right| = \left| \text{Tr}(Q) \right| = \left| \text{Tr}(M'^{1/2}(I - M)M'^{1/2}) \right|
\]

\[
= \left| \text{Tr}(M'(I - M)) \right| = \left| \text{Tr}((I - (I - M'))(I - M)) \right|
\]

\[
\leq \left| \text{Tr}(I - M) \right| + \left| \text{Tr}((I - M')(I - M)) \right| \quad \text{(28)}
\]

\[
= \gamma + \left| \text{Tr}((I - M')(I - M)) \right| \quad \text{(since \( \left| \text{Tr}(I - M) \right| \leq \gamma \) for all \( S \in E(\gamma) \))}
\]

\[
\leq \gamma + ||M' ||_F \cdot ||M - I||_F \quad \text{(by von Neumann and Cauchy-Schwarz inequalities)}
\]

\[
\leq \gamma + \gamma^2
\]

We thus have

\[
Y^T M'^{1/2}(I - M)M'^{1/2}Y = \sum_{i=1}^{d} (1 - \tilde{\lambda}_i)Z_i^2
\]

\[
= \sum_{i=1}^{d} (1 - \tilde{\lambda}_i) + \sum_{i=1}^{d} (1 - \tilde{\lambda}_i)(Z_i^2 - 1)
\] (29)

We now use a calculation analogous to the above for (1) to show that \( \left| \sum_{i=1}^{d} (1 - \tilde{\lambda}_i)(Z_i^2 - 1) \right| \leq \frac{1}{2} \cdot \frac{1}{100} \) with probability at least \( 1 - n^{-40}/4 \). Indeed, we first verify that the variance is bounded by

\[
O\left( \sum_{i=1}^{d} (1 - \tilde{\lambda}_i)^2 \right) = O(||Q||_F^2)
\]

\[
= O(||M'^{1/2}(I - M)M'^{1/2}||_F^2)
\]

\[
\leq O( ||M'||_2^2 ||I - M||_F^2) \quad \text{(by sub-multiplicativity)}
\]

\[
\leq O( ||I||_2 + ||M' - I||_F^2 ) ||I - M||_F^2
\]

\[
\leq O( ||I - M||_F^2 )
\]

\[
= O(\gamma^2). \quad \text{(30)}
\]

We also have

\[
|(1 - \tilde{\lambda}_i)Y_i| \leq ||Q||_F C \sqrt{\log n}
\]

\[
\leq ||M'||_2 ||I - M||_F C \sqrt{\log n} \quad \text{(by sub-multiplicativity)}
\]

\[
\leq ( ||I||_2 + ||M' - I||_F^2 ) ||I - M||_F C \sqrt{\log n}
\]

\[
\leq 2 ||I - M||_F C \sqrt{\log n}
\]

\[
\leq 2 \gamma \cdot C \sqrt{\log n}
\]

for all \( i \) with probability at least \( 1 - 1/n^{40}/5 \) as long as \( C > 0 \) is larger than an absolute constant. We thus have by Theorem 24 (applied to clipped variables and then unclipping by a union bound as in (1)) for all \( t \geq 0 \) that

\[
\Pr[|Y^T M'^{1/2}(I - M)M'^{1/2}Y - \sum_{i=1}^{d} (1 - \tilde{\lambda}_i)| > t] < \exp\left( -O\left( \frac{1}{2} t^2 \cdot \frac{1}{O(\sum_{i=1}^{d} (1 - \tilde{\lambda}_i)^2) + (\frac{1}{2} \gamma C \sqrt{\log n} t)} \right) \right) + n^{-40}/5.
\]
Setting $t = \frac{1}{2} \frac{1}{100}$, and using the upper bound $O(\sum_i (1 - \bar{\lambda}_i)^2) = O(\gamma^2)$ obtained in (30), we get

$$
\Pr[\|Y^TM^{1/2}(I - M)M^{1/2}Y - \sum_{i=1}^d (1 - \bar{\lambda}_i)\| > \frac{1}{2} \frac{1}{100}] < \exp \left( - \frac{1}{2} \frac{1}{100} \right) \leq \exp(-\Omega(1/(\gamma \sqrt{\log n})) + n^{-40}/5
$$

since $\gamma \leq 1/\log^2 n$ by assumption, for a sufficiently large $n$. Since $|\sum_{i=1}^d (1 - \bar{\lambda}_i)| \leq \gamma + 2\gamma^2 \leq \frac{1}{2} \frac{1}{100}$ by (28), we get by triangle inequality that

$$
\Pr_{X \sim N(0,M')}[\|X^T(I - M)X\| > \frac{1}{100}] \leq n^{-40}/4.
$$

Similarly to (1) above, we have, when $X \sim N(0,M'), X = M^{1/2}Y, Y \sim N(0,I_d)$,

$$
X^T(I - M)^2X = Y^TM^{1/2}(I - M)^2M^{1/2}Y = \sum_{i=1}^d \tilde{\tau}_i Z_i^2
$$

$$
\leq \text{Tr}(M^{1/2}(I - M)^2M^{1/2}) \cdot \max_{i \in [d]} Z_i^2
$$

$$
\leq O(\log n) \cdot \text{Tr}(M^{1/2}(I - M)^2M^{1/2})
$$

with probability at least $1 - n^{-40}/2$ over the choice of $X$, as $\max_{i \in [d]} Z_i^2 \leq C \log n$ with high probability if $C$ is a sufficiently large constant by standard properties of Gaussian random variables. Since $\text{Tr}(M^{1/2}(I - M)^2M^{1/2}) = \text{Tr}(M'(I - M)^2) \leq 2\|I - M\|_F^2$ (as $\gamma < 1/\log^2 n < 1/3$ by assumption of the lemma), we get

$$
X^T(I - M)^2X \leq O(\log n) \cdot \text{Tr}(M^{1/2}(I - M)^2M^{1/2}) \leq O(\log n) \cdot \gamma^2 \leq \frac{1}{100} (\text{since } \gamma < 1/\log^2 n)
$$

with probability at least $1 - n^{-40}/4$. A union bound over the failure events yields $\Pr_{X \sim N(0,M')}[X \notin \mathcal{T}(S,U)] < n^{-40}$, as required.

This completes the proof.

**Proof of Lemma 20**: By assumption that $S \in \mathcal{E}(\gamma)$ we have that $\|I - M\|_2 \leq \gamma$, so Taylor expansion is valid and gives

$$
\frac{1}{2} x^T x - \frac{1}{2} x^T M^{-1} x - \frac{1}{2} \log \det M = \frac{1}{2} x^T(I - M)x + \frac{1}{2} \text{Tr}(I - M) + R(x),
$$

where for all $x \in \mathcal{T}(S,U)$ one has $R(x) \leq \sum_{k \geq 2} x^T(I - M)^2x + \text{Tr}(I - M)^k$.

We have by Lemma 18 that $R(x) \leq C(x^T(I - M)^2x + \|I - M\|_F^2)$ for an absolute constant $C > 0$, for all $x \in \mathcal{T}(S,U)$ and $S \in \mathcal{E}(\gamma)$. We thus have

$$
e^{-\frac{1}{2} x^T(I - M)x + \frac{1}{2} \text{Tr}(I - M) - C(x^T(I - M)^2x + \|I - M\|_F^2)}
$$

$$\leq e^{-\frac{1}{2} x^T x + \frac{1}{2} \text{Tr}(I - M) - \frac{1}{2} x^T M^{-1} x - \frac{1}{2} \log \det M}
$$

$$\leq e^{-\frac{1}{2} x^T(I - M)x + \frac{1}{2} \text{Tr}(I - M) + C(x^T(I - M)^2x + \|I - M\|_F^2)}
$$

for all such $M$ and $x$.

We now Taylor expand $e^{-\frac{1}{2} x^T(I - M)x + \frac{1}{2} \text{Tr}(I - M) + A(x^T(I - M)^2x + \|I - M\|_F^2)}$, where $A$ is any constant (positive or negative), getting

$$
e^{-\frac{1}{2} x^T(I - M)x + \frac{1}{2} \text{Tr}(I - M) + A(x^T(I - M)^2x + \|I - M\|_F^2)}
$$

$$= \sum_{k \geq 1} \left(-\frac{1}{2} x^T(I - M)x + \frac{1}{2} \text{Tr}(I - M) + A(x^T(I - M)^2x + \|I - M\|_F^2)\right)^k / k!.
$$

(32)
For $k = 2$ we have

$$\left| \left( -\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) + x^T (I - M)^2 x + \|I - M\|^2_F \right) \right|^2 + \frac{1}{8} x^T (I - M)x \cdot \text{Tr}(I - M) \right) \leq C \left( (x^T (I - M)x)^2 + (\text{Tr}(I - M))^2 + x^T (I - M)^2 x + ||I - M||^2_F \right),$$

where we used the fact $|x^T (I - M)x| \leq \frac{1}{100}$ for $x \in \mathcal{T}\{S, U\}$ and $|\text{Tr}(I - M)| \leq \gamma \leq \frac{1}{100}$ for $S \in \mathcal{E}(\gamma)$. For all $k \geq 3$ we use the bound

$$\left| \left( -\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) + x^T (I - M)^2 x + ||I - M||^2_F \right) \right|^k \leq \left( |x^T (I - M)x| + \frac{1}{2} |\text{Tr}(I - M)| + x^T (I - M)^2 x + ||I - M||^2_F \right)^{k}$$

which we used to go from the second line to the third, and the last line follows from the observation that every term in the expansion of $\left( |x^T (I - M)x| + \frac{1}{2} |\text{Tr}(I - M)| + x^T (I - M)^2 x + ||I - M||^2_F \right)^{3}$ contains either at least a square of one of the first two terms or at least one of the last two.

Substituting these bounds into (33), we get

$$e^{-\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) + A(x^T (I - M)^2 x + ||I - M||^2_F)}$$

$$= \sum_{k \geq 1} \left( -\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) + A(x^T (I - M)^2 x + ||I - M||^2_F) \right)^k /k!$$

$$\leq -\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) - \frac{1}{8} x^T (I - M)x \cdot \text{Tr}(I - M)$$

$$+ C \left( |x^T (I - M)x|^2 + x^T (I - M)^2 x + \text{Tr}(I - M)^2 + ||I - M||^2_F \right) \quad \text{(for a constant } C > 0 \text{ that may depend on } A)$$

$$+ \sum_{k \geq 1} (A + 1)^k \left( |x^T (I - M)x|^2 + x^T (I - M)^2 x + ||I - M||^2_F \right)/k!$$

$$\leq -\frac{1}{2} x^T (I - M)x + \frac{1}{2} \text{Tr}(I - M) + C''(x^T (I - M)^2 x + \text{Tr}(I - M)^2 + x^T (I - M)x^2 + ||I - M||^2_F)$$

for an absolute constant $C'' > 0$. The provides the upper bound in the claimed result. The lower bound is provided by a similar calculation, which we omit.

**Proof of Lemma 23.** Since $\mathbb{E}|(Z - 1)^2| \leq \epsilon$ by assumption of the lemma, for any event $\mathcal{E}$ one has $\mathbb{E}|(Z - 1)^2 \cdot I_{\mathcal{E}}| \leq \epsilon$, where $I_{\mathcal{E}}$ is the indicator of $\mathcal{E}$, the complement of $\mathcal{E}$. This also means that

$$\mathbb{E}|(Z - 1)^2| \leq \epsilon / \mathbb{P}[\mathcal{E}].$$

On the other hand, by Jensen’s inequality

$$\mathbb{E}|Z - 1||\mathcal{E}| \leq (\mathbb{E}|(Z - 1)^2| \mathcal{E}|)^{1/2},$$

and putting these two bounds together we get

$$\mathbb{E}|Z - 1| \cdot I[\mathcal{E}] = \mathbb{E}|Z - 1||\mathcal{E}| \cdot \mathbb{P}[\mathcal{E}] \leq \mathbb{P}[\mathcal{E}] \cdot (\mathbb{E}|(Z - 1)^2| \mathcal{E}|)^{1/2} \leq \mathbb{P}[\mathcal{E}] \cdot (\epsilon / \mathbb{P}[\mathcal{E}])^{1/2} = \sqrt{\epsilon \cdot \mathbb{P}[\mathcal{E}]}.$$
This means that
\[
|E[Z] - E[Z|\mathcal{E}]| \leq |E[Z] - \frac{1}{Pr[\mathcal{E}]}E[Z \cdot I_{\mathcal{E}}]|
\]
\[
\leq |E[Z] - \frac{1}{Pr[\mathcal{E}]}E[Z] + \frac{1}{Pr[\mathcal{E}]}E[Z \cdot I_{\mathcal{E}}]|
\]
\[
\leq E[Z] \left( \frac{1}{1 - Pr[\mathcal{E}]} - 1 \right) + \frac{1}{Pr[\mathcal{E}]}E[Z \cdot I_{\mathcal{E}}]
\]
\[
\leq E[Z] \cdot 2Pr[\mathcal{E}] + 2E[Z \cdot I_{\mathcal{E}}] \quad \text{(since } \frac{1}{1 - x} - 1 \leq 2x \text{ for } x \in (0, 1/2))
\]
\[
\leq E[Z] \cdot 2Pr[\mathcal{E}] + 2(Pr[\mathcal{E}] + E[|Z - 1| \cdot I_{\mathcal{E}}])
\]
\[
\leq 2(1 + E[Z])Pr[\mathcal{E}] + 2\sqrt{\epsilon}Pr[\mathcal{E}].
\]

\[
\frac{\sqrt{\epsilon}}{Pr[\mathcal{E}]}.
\]

\[\Box\]

A.2 Proofs of moment bounds (Lemma 21 and Lemma 22)

Proof of Lemma 21 and Lemma 22: We start by noting that for every \(i, j \in [1 : d]\) the matrix \(M = U^T S^T SU\) satisfies

\[
M_{ij} = \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{b=1}^{n} S_{r,a} U_{a,i} S_{r,b} U_{b,j}
\]
\[
= \sum_{a=1}^{n} U_{a,i} U_{a,j} \left( \sum_{r=1}^{B} S_{r,a}^2 \right) + \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{b=1}^{n} S_{r,a} U_{a,i} S_{r,b} U_{b,j}
\]
\[
= \delta_{i,j} + \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{b=1, a \neq b}^{n} S_{r,a} U_{a,i} S_{r,b} U_{b,j},
\]

where \(\delta_{i,j}\) equals 1 if \(i = j\) and equals 0 otherwise. We thus have, for every \(i, j \in [1 : d]\), that

\[
(M - I)_{ij} = \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{b=1, a \neq b}^{n} S_{r,a} U_{a,i} S_{r,b} U_{b,j},
\]

which in particular means that

\[
\text{Tr}(I - M) = -\sum_{i} (M - I)_{ii} = -\sum_{i} \sum_{r=1}^{B} \sum_{a=1, a \neq b}^{n} S_{r,a} U_{a,i} S_{r,b} U_{b,i},
\]
\[
= -\sum_{r=1}^{B} \sum_{a=1, a \neq b}^{n} S_{r,a} S_{r,b} \cdot U_a U_b^T,
\]

(note that it immediately follows that \(E_S[\text{Tr}(I - M)] = 0\), as \(E_S[S_{r,a} S_{r,b}] = 0\) for \(a \neq b\)) and

\[
x^T (I - M)x = -\sum_{i,j} (M - I)_{ij} x_i x_j = -\sum_{i,j} \sum_{r=1}^{B} \sum_{a=1, a \neq b}^{n} S_{r,a} U_{a,i} S_{r,b} U_{b,j} x_i x_j
\]
\[
= -\sum_{r=1}^{B} \sum_{a=1, a \neq b}^{n} S_{r,a} S_{r,b} (U x)_a (U x)_b.
\]
(note that it immediately follows that $E_S[x^T(I - M)x] = 0$ for all $x$, as $E_S[S_{r,a}S_{r,b}] = 0$ for $a \neq b$).

We also have

$$(M - I)_{ij}^2 = \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{r'=1}^{B} \sum_{a'=1}^{n} S_{r,a}U_{a,i}S_{r,b}U_{b,j}S_{r',c}U_{c,i}S_{r',d}U_{d,j}$$

and hence

$$||I - M||^2_F = \sum_{ij} (M - I)_{ij}^2 = \sum_{ij} \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{r'=1}^{B} \sum_{a'=1}^{n} S_{r,a}U_{a,i}S_{r,b}U_{b,j}S_{r',c}U_{c,i}S_{r',d}U_{d,j}$$

$$= \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{r'=1}^{B} \sum_{a'=1}^{n} S_{r,a}S_{r,b}S_{r',c}S_{r',d}(\sum_{i} U_{a,i}U_{c,i})(\sum_{j} U_{b,j}U_{d,j})$$

$$= \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{r'=1}^{B} \sum_{a'=1}^{n} S_{r,a}S_{r,b}S_{r',c}S_{r',d} \cdot U_aU_c^T \cdot U_bU_d^T$$

$$= \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{r'=1}^{B} \sum_{a'=1}^{n} S_{r,a}S_{r,b}S_{r',c}S_{r',d} \cdot U_aU_b^T \cdot U_aU_b^T$$

We also need

$$x^T(I - M)^2x = ||(I - M)x||_2^2 = \sum_{i=1}^{d} \left( \sum_{j=1}^{d} (I - M)_{ij}x_j \right)^2$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} x_j^2 \cdot \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{a'=1}^{n} S_{r,a}U_{a,i}S_{r,b}U_{b,j} \cdot U_{a,i}U_{a,i}^T \cdot U_{b,j}U_{b,j}$$

$$= \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{a'=1}^{n} S_{r,a}S_{r,b}S_{r',c}S_{r',d} \cdot \sum_{i=1}^{d} U_{a,i}U_{a,i}^T \cdot \sum_{j=1}^{d} U_{b,j}U_{b,j}$$

$$= \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{a'=1}^{n} S_{r,a}S_{r,b}S_{r',c}S_{r',d} \cdot U_aU_b^T \cdot (U_aU_b^T)^T$$

$$= \sum_{r=1}^{B} \sum_{a=1}^{n} \sum_{a'=1}^{n} S_{r,a}S_{r,b}S_{r',c}S_{r',d} \cdot U_aU_b^T \cdot (U_x)_{b,a}(U_x)_{b,a}$$

Bounding $E_S[||I - M||^2_F], E_S[<x^T(I - M)x>^2], E_S[x^T(I - M)^2x], E_S[(x^T(I - M)x)Tr(I - M)], E_S[Tr(I - M)^2]

We first note that for for any $r_1, r_2$ and $a \neq b$, $a \neq b$, the quantity

$$E_S[S_{r_1,a}S_{r_1,b}S_{r_2,a}S_{r_2,b}]$$

is only nonzero when $r_1 = r_2$ and $\{a_1, b_1, a_2, b_2\}$ contains two distinct elements, each with multiplicity 2 (let $L_1(\{a_1, b_1\})$ denote the indicator of the latter condition). In that case one has $E_S[S_{r_1,a}S_{r_1,b}S_{r_2,a}S_{r_2,b}] =$
We have $r \cdot |U| \leq 1$ and $A + B + C + D + E + F + G + H = 2$. We thus have

$$\left| \sum_{r=1}^{B} \sum_{r_2=1}^{B} \sum_{a_1, b_1=1}^{n} \sum_{a_2, b_2=1}^{n} \mathbf{E}_S[S_{r_1, a_1} S_{r_2, b_2} S_{r_2, a_2} S_{r_2, b_2}] \cdot (U_{a_1, b_1}^T A (U_{b_2, a_2}^T)^B \cdot ((U_{x_1})_{a_1} (U_{b_2})_{a_2})^C ((U_{x_2})_{b_1} (U_{x_2})_{b_2})^D \cdot ((U_{x_1})_{a_1} (U_{b_2})_{a_2})^E (U_{a_1, b_1}^T)^F \cdot ((U_{x_2})_{a_2} (U_{b_2})_{b_2})^G (U_{a_2, b_2}^T)^H, \right|$$

where $A, B, C, D, E, F, G, H \in \{0, 1\}$ and $A + B + C + D + E + F + G + H = 2$. We thus have

$$\left| \sum_{r=1}^{B} \sum_{r_2=1}^{B} \sum_{a_1, b_1=1}^{n} \sum_{a_2, b_2=1}^{n} \mathbf{E}_S[S_{r_1, a_1} S_{r_2, b_2} S_{r_2, a_2} S_{r_2, b_2}] \cdot (U_{a_1, b_1}^T A (U_{b_2, a_2}^T)^B \cdot ((U_{x_1})_{a_1} (U_{b_2})_{a_2})^C ((U_{x_2})_{b_1} (U_{x_2})_{b_2})^D \cdot ((U_{x_1})_{a_1} (U_{b_2})_{a_2})^E (U_{a_1, b_1}^T)^F \cdot ((U_{x_2})_{a_2} (U_{b_2})_{b_2})^G (U_{a_2, b_2}^T)^H, \right| \leq \frac{1}{B} \sum_{a_1, b_1=1}^{n} \sum_{a_2, b_2=1}^{n} \mathbf{I}_s(\{a_q, b_q\}_{q=1}^{Q}) |U_{a_1}^T|^A |U_{b_2, a_2}|^B \cdot ((U_{x_1})_{a_1} (U_{b_2})_{a_2})^C ((U_{x_1})_{b_1} (U_{x_2})_{b_2})^D \cdot ((U_{x_1})_{a_1} (U_{b_2})_{b_2})^E (U_{a_2, b_2}^T)^F \cdot ((U_{x_2})_{a_2} (U_{b_2})_{b_2})^G (U_{a_2, b_2}^T)^H.$$
Bounding \( \mathbb{E}_S[(x^T(I-M)x)^2 \text{Tr}(I-M)] \), \( \mathbb{E}_S[x^T(I-M)^2 x \cdot \text{Tr}(I-M)] \), \( \mathbb{E}_S(||I-M||^2_p \cdot \text{Tr}(I-M)] \), \( \mathbb{E}_S[(x^T(I-M)x)^2 \cdot x^T(I-M)x] \). All of the above expressions can be written as

\[
\sum_{r_1=1}^{B} \sum_{r_2=1}^{B} \sum_{r_3=1}^{B} \sum_{a_1=1}^{n} \sum_{a_2=1}^{n} \sum_{a_3=1}^{n} \sum_{a_4=1}^{n} \sum_{b_1=1}^{n} \sum_{b_2=1}^{n} \sum_{b_3=1}^{n} \mathbb{E}_S[S_{r_1,a_1}S_{r_1,b_1}S_{r_2,a_2}S_{r_2,b_2}S_{r_3,a_3}S_{r_3,b_3}]
\cdot (U_{1a_1}U_{1b_1}^T)A(U_{1b_1}U_{2b_2}^T)B \cdot ((U_{2a_1}(U_{x}a_2)C((U_{x}a_1(U_{x}b_1))D \cdot ((U_{x}a_1(U_{x}b_1))E(U_{1a_1}U_{1b_1}^T)F \cdot ((U_{x}a_2(U_{x}b_2))G(U_{2a_2}U_{2b_2}^T)^H
\cdot ((U_{x}a_3(U_{x}b_3))^I(U_{3a_3}U_{3b_3})^J
\]

where \( A, B, C \ldots \) are in \( \{0, 1\} \) and \( A + B + C + D + E + F + G + H + I + J = 3 \).

We first note that for for any \( r_1, r_2, r_3 \) and \( a_1 \neq b_1, a_2 \neq b_2, a_3 \neq b_3 \) the quantity

\[
\mathbb{E}_S[S_{r_1,a_1}S_{r_1,b_1}S_{r_2,a_2}S_{r_2,b_2}S_{r_3,a_3}S_{r_3,b_3}]
\]

is only nonzero when \( r_1 = r_2 = r_3 \) and \( \{a_1, b_1, a_2, b_2, a_3, b_3\} \) contains three distinct elements, each with multiplicity 2. Let \( \text{I}_r(\{a_q, b_q\}_{q=1}^{3}) \) denote the indicator of the latter condition. In that case one has \( \mathbb{E}_S[S_{r_1,a_1}S_{r_1,b_1}S_{r_2,a_2}S_{r_2,b_2}S_{r_3,a_3}S_{r_3,b_3}] = 1/B^3 \). Note we cannot have \( a_1 = a_2 = a_3 \) and \( b_1 = b_2 = b_3 \) since the expectation is 0 in that case.

Similarly to the above, it thus suffices to bound

\[
\frac{1}{B^2} \sum_{a_1 \neq b_1}^{n} \sum_{a_2 \neq b_2}^{n} \sum_{a_3 \neq b_3}^{n} \text{I}_r(\{a_q, b_q\}_{q=1}^{3}) \cdot ((U_{a_1}U_{a_2}^T)A(U_{b_1}U_{b_2}^T)B \cdot ((U_{x}a_1(U_{x}a_2)C((U_{x}b_1(U_{x}b_2))D \cdot ((U_{x}a_1(U_{x}b_1))E(U_{1a_1}U_{1b_1}^T)F \cdot ((U_{x}a_2(U_{x}b_2))G(U_{2a_2}U_{2b_2}^T)^H
\cdot ((U_{x}a_3(U_{x}b_3))^I(U_{3a_3}U_{3b_3})^J
\]

where we used Cauchy-Schwarz and the assumption that \( x \in \mathcal{T}^* \) (and hence \( x \) is not correlated with any of the rows of \( U \) too much), as above.

Since we are only summing over \( \{a_1, a_2, a_3, b_1, b_2, b_3\} \) that contain three distinct elements, the expression above is upper bounded by

\[
(O(\log n))^{C+D+E+G+I} \frac{1}{B^2} \sum_{a,c,b}^{n} ||U_{a}||^2 ||U_{b}||^2 ||U_{c}||^2
\]

where we used the fact that \( \sum_a ||U_a||^2 = d \) and that in all cases, \( C + D + E + G + I \leq 2 \).
Bounding $E_S[(x^T(I-M)x)^2 + x^T(I-M)^2x + ||I-M||_F^2 + (\text{Tr}(I-M))^2]^2]$. All of the pairwise products arising in the expansion of the above expressions can be written as

$$
\sum_{r_1=1}^B \sum_{r_2=1}^B \sum_{r_3=1}^B \sum_{n=1}^n \sum_{n=1}^n \sum_{n=1}^n \sum_{n=1}^n E_S[S_{r_1,a_1}S_{r_1,b_1}S_{r_2,a_2}S_{r_2,b_2}S_{r_3,a_3}S_{r_3,b_3}S_{r_4,a_4}S_{r_4,b_4}]
$$

$$
\cdot (U_{a_1}U_{a_2})^A(U_{b_1}U_{b_2})^B \cdot ((U_x)_{a_1}(U_x)_{a_2})^C((U_x)_{b_1}(U_x)_{b_2})^D \cdot ((U_x)_{a_3}(U_x)_{b_3})^E(U_{a_1}U_{b_1})^F \cdot ((U_x)_{a_2}(U_x)_{b_2})^G(U_{a_2}U_{b_2})^H
$$

$$
\cdot (U_{a_3}U_{a_4})^A(U_{b_3}U_{b_4})^B' \cdot ((U_x)_{a_3}(U_x)_{a_4})^C'((U_x)_{b_3}(U_x)_{b_4})^D' \cdot ((U_x)_{a_4}(U_x)_{b_4})^E'(U_{a_3}U_{b_3})^F' \cdot ((U_x)_{a_4}(U_x)_{b_4})^G'(U_{a_4}U_{b_4})^H',
$$

where $A, B, C, D, E, F, G, H, A', B', C', D', E', F', G', H' \in \{0, 1\}$ and add up to 4.

We now need to consider two cases.

**Case 1:** the number of distinct elements in $\{a_1, a_2, b_2, a_3, b_3, a_4, b_4\}$ is four, each occurring with multiplicity 2 (let $I_q(\{a_q, b_q\}^4_{q=1})$ denote the indicator of the latter condition). Then

$$
E_S[S_{r_1,a_1}S_{r_1,b_1}S_{r_2,a_2}S_{r_2,b_2}S_{r_3,a_3}S_{r_3,b_3}S_{r_4,a_4}S_{r_4,b_4}]
$$

contributes $1/B^4$. In this case the number of distinct elements in $\{r_1, r_2, r_3, r_4\}$ cannot be larger than 2.

It thus suffices to bound

$$
\frac{1}{B^2} \sum_{a_1 \neq a_2}^n \sum_{a_3 = a_1}^n \sum_{a_4 = a_2}^n \sum_{r_1 = 1}^B \sum_{a_1 = 1}^n \sum_{a_2 = 1}^n \sum_{a_3 = 1}^n \sum_{a_4 = 1}^n I_q(\{a_q, b_q\}^4_{q=1})
$$

$$
\cdot (|U_{a_1}|_2||U_{a_2}|_2)^A(||U_{b_1}|_2||U_{b_2}|_2)^B \cdot (||U||_1||U||_1)^C(||U_{b_1}|_2||U_{b_2}|_2)^D \cdot (||U_{a_1}|_2||U_{b_1}|_2)^E(||U_{a_1}|_2||U_{b_1}|_2)^F
$$

$$
\cdot (||U_{a_2}|_2||U_{b_2}|_2)^G(||U_{a_2}|_2||U_{b_2}|_2)^H
$$

$$
\cdot (||U_{a_3}|_2||U_{a_4}|_2)^A(||U_{b_3}|_2||U_{b_4}|_2)^B' \cdot (||U_{a_3}|_2||U_{a_4}|_2)^C'(||U_{b_3}|_2||U_{b_4}|_2)^D' \cdot (||U_{a_3}|_2||U_{b_3}|_2)^E'(||U_{a_3}|_2||U_{b_3}|_2)^F'
$$

$$
\cdot (||U_{a_4}|_2||U_{b_4}|_2)^G'(||U_{a_4}|_2||U_{b_4}|_2)^H'
$$

where we used Cauchy-Schwarz and the assumption that $x \in \mathcal{T}^*$ (and hence $x$ is not correlated with any of the rows of $U$ too much), as above.

Since we are only summing over $\{a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4\}$ that contain three distinct elements, each of multiplicity two, the expression above is upper bounded by

$$
(O(\log n))^2 \frac{1}{B^2} \sum_{a,b,c,d}^n ||U_a||_2^2||U_b||_2^2||U_c||_2^2||U_d||_2^2
$$

$$
\leq (O(\log n))^2 \frac{d^4}{B^2}
$$

where we used the fact that $\sum ||U_a||_2^4 = d$.  

22
Case 2: the number of distinct elements in \( \{a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4\} \) is two, each occurring with multiplicity 4 (let \( I, \{a_q, b_q\}_{q=1}^4 \) denote the indicator of the latter condition) Then

\[
E_S[S_{r_1, a_1} S_{r_1, b_1} S_{r_2, a_2} S_{r_2, b_2} S_{r_3, a_3} S_{r_3, b_3} S_{r_4, a_4} S_{r_4, b_4}]
\]

contributes \( 1/B^2 \). In this case the number of distinct elements in \( \{r_1, r_2, r_3, r_4\} \) has to be one, since each column of \( S \) has a single non-zero entry and necessarily \( a_1 = a_2 = a_3 = a_4 \) and \( b_1 = b_2 = b_3 = b_4 \).

It thus suffices to bound

\[
\frac{1}{B} \sum_{a_1, b_1=1}^n \sum_{a_2, b_2=1}^n \sum_{a_3, b_3=1}^n \sum_{a_4, b_4=1}^n I_r(\{a_q, b_q\}_{q=1}^4).
\]

\[
\cdot (U_{a_1} U_{a_2}^T) A (U_{b_1} U_{b_2}^T)^B \cdot ( (U_x)_{a_1} (U_x)_{a_2} ) C ((U_x)_{b_1} (U_x)_{b_2})^D \cdot ( (U_x)_{a_1} (U_x)_{b_1} )^E (U_{a_1} U_{b_1}^T)^F \cdot ( (U_x)_{a_2} (U_x)_{b_2} )^G (U_{a_2} U_{b_2}^T)^H
\]

\[
\cdot (U_{a_3} U_{a_4}^T) A' (U_{b_3} U_{b_4}^T)^B' \cdot ( (U_x)_{a_3} (U_x)_{a_4} ) C' ((U_x)_{b_3} (U_x)_{b_4})^D' \cdot ( (U_x)_{a_3} (U_x)_{b_3} )^E' (U_{a_3} U_{b_3}^T)^F' \cdot ( (U_x)_{a_4} (U_x)_{b_4} )^G' (U_{a_4} U_{b_4}^T)^H'
\]

\[
\leq (O(\log n))^2 \frac{1}{B} \sum_{a_1, b_1=1}^n \sum_{a_2, b_2=1}^n \sum_{a_3, b_3=1}^n \sum_{a_4, b_4=1}^n I_r(\{a_q, b_q\}_{q=1}^4).
\]

\[
\cdot (||U_{a_1}||_2 ||U_{a_2}||_2)^A (||U_{b_1}||_2 ||U_{b_2}||_2)^B \cdot (||U||_a ||U||_a') C (||U_{b_1}||_2 ||U_{b_2}||_2)^D \cdot (||U_{a_1}||_2 ||U_{b_1}||_2)^E (||U_{a_1}||_2 ||U_{b_1}||_2)^F
\]

\[
\cdot (||U_{a_2}||_2 ||U_{b_2}||_2)^G (||U_{a_2}||_2 ||U_{b_2}||_2)^H
\]

\[
\cdot (||U_{a_3}||_2 ||U_{a_4}||_2)^A' (||U_{b_3}||_2 ||U_{b_4}||_2)^B' \cdot (||U_{a_3}||_2 ||U_{a_4}||_2)^C' (||U_{b_3}||_2 ||U_{b_4}||_2)^D' \cdot (||U_{a_3}||_2 ||U_{b_3}||_2)^E' (||U_{a_3}||_2 ||U_{b_3}||_2)^F'
\]

\[
\cdot (||U_{a_4}||_2 ||U_{b_4}||_2)^G' (||U_{a_4}||_2 ||U_{b_4}||_2)^H'
\]

where we used Cauchy-Schwarz and the assumption that \( x \in T^* \) (and hence \( x \) is not correlated with any of the rows of \( T \) too much), as above.

Since we are only summing over \( \{a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4\} \) that contain two distinct elements, each of multiplicity four, the expression above is upper bounded by

\[
(O(\log n))^2 \frac{1}{B} \sum_{a, b} ||U_a||_2^2 ||U_b||_2^2
\]

\[
= (O(\log n))^2 \frac{1}{B} \sum_{a, b} ||U_a||_2^2 ||U_b||_2^2 \quad \text{ (since } ||U_a||_2 \leq 1 \text{ for all } a) \]

\[
\leq (O(\log n))^2 d^2 \frac{1}{B},
\]

where we used the fact that \( \sum_a ||U_a||_2^2 = d \). \( \square \)