RANDOMIZED ALGORITHMS FOR ESTIMATING THE TRACE OF AN IMPLICIT SYMMETRIC POSITIVE SEMI-DEFINITE MATRIX

HAIM AVRON AND SIVAN TOLEDO

Abstract. We analyze the convergence of randomized trace estimators. Starting at 1989, several algorithms have been proposed for estimating the trace of a matrix by $\frac{1}{M}\sum_{i=1}^{M} z_i^T A z_i$, where the z_i are random vectors, have been proposed; different estimators use different distributions for the z_is, all of which lead to $E(\frac{1}{M}\sum_{i=1}^{M} z_i^T A z_i) = \mathrm{trace}(A)$. These algorithms are useful in applications in which there is no explicit representation of A but rather an efficient method compute $z^T A z$ given z. Existing results only analyze the variance of the different estimators. In contrast, we analyze the number of samples M required to guarantee that with probability at least $1-\delta$, the relative error in the estimate is at most ϵ. We argue that such bounds are much more useful in applications than the variance. We found that these bounds rank the estimators differently than the variance; this suggests that minimum-variance estimators may not be the best.

We also make two additional contributions to this area. The first is a specialized bound for projection matrices, whose trace (rank) needs to be computed in electronic structure calculations. The second is a new estimator that uses less randomness than all the existing estimators.

1. Introduction

Finding the trace of an explicit matrix is a simple operation. But there are application areas where one needs to compute the trace of an implicit matrix, that is, a matrix represented as a function. For example, in lattice Quantum Chromodynamics, one often needs to compute the trace of a function of a large matrix, $\mathrm{trace}(f(A))$. Explicitly computing $f(A)$ for large matrices is not practical, but computing the bilinear form $x^T f(A) x$ for an arbitrary x is feasible [5, 4]. Other examples include the regularized solution of least-squares problems using the Generalized Cross-Validation approach (see [9]) and computing the number of triangles in a graph [14].

The standard approach for computing the trace of an implicit function is Monte-Carlo simulation, where the trace is estimated by $\frac{1}{M}\sum_{i=1}^{M} z_i^T A z_i$, where the z_i are random vectors. The original method is due to Hutchinson [9]. Although this method has been improved over the years ([6, 10, 16]), no paper to date has presented a theoretical bound on the number of samples required to achieve an ϵ-approximation of the trace; only the variance of estimators has been analyzed.

This paper makes four significant contributions to this area:

(1) We provide rigorous bounds on the number of Monte-Carlo samples required to achieve a maximum error ϵ with probability at least $1-\delta$ in several trace estimators. The bounds are surprising: the method with the best bound is not the method with the smallest variance.
(2) We provide specialized bounds for the case of projection matrices, which are important in certain applications.
(3) We propose a new trace estimator in which the z_is are random columns of a unitary matrix with entries that are small in magnitude. This estimator converges slower than known ones, but it also uses fewer random bits.

Date: April, 2010.

(4) We experimentally evaluate the convergence of the three methods on a few interesting matrices.

2. Hutchinson's Method and Related Work

The standard Monte-Carlo method for estimating the trace of an implicit method is due to Hutchinson [9], who proves the following Lemma.

Lemma 1. *Let A be an $n \times n$ symmetric matrix with* $\mathrm{trace}(A) \neq 0$. *Let z be a random vector whose entries are i.i.d Rademacher random variables (*$\Pr(z_i = \pm 1) = 1/2$*). $z^T Az$ is an unbiased estimator of* $\mathrm{trace}(A)$ *i.e.,*

$$\mathrm{E}(z^T Az) = \mathrm{trace}(A)$$

and

$$\mathrm{Var}(z^T Az) = 2\left(\|A\|_F^2 - \sum_{i=1}^{n} A_{ii}^2\right).$$

If we examine the variance term we see that intuitively it measures how much of the matrix's "energy" (i.e., the Frobenius norm) is on the diagonal. It is easy to see that for a general matrix Hutchinson's method can be ineffective because the variance can be arbitrarily large. Even for a symmetric positive definite the variance can be large: the variance for the matrix of all 1's, which is symmetric semi-definite, is $2(n^2 - n)$, whereas the trace is only n. This matrix can be perturbed to definiteness without a significant impact on the trace or variance. Such a large variance precludes the use of Chebyshev's inequality to bound the number of iterations required to obtain a given relative error in the trace. For such a bound to hold, the variance must be $o(\mathrm{trace}(A)^2)$.

Lemma 1 does not give a rigorous bound on the number of samples/matrix multiplications. This difficulty carries over to applications of this method, such as [5, 4]. Hutchinson's method has been improved over the years, but the improvements do not appear to have addressed this issue. Wong et al. [16] suggest using test vectors z that are derived from columns of an Hadamard matrix. Iitaka and Ebisuzaki [10] generalized Hutchinson's estimator by using *complex* i.i.d's with unit magnitude; they showed that the resulting estimator has lower variance than Hutchison's (but the computation cost is also higher). Silver and Röder [13] use Gaussian i.i.d variables, but without any analysis. Bekas et al. [6] focus on approximating the actual diagonal values, also using vectors derived from an Hadamard matrix.

In Section 7 below we show that it is possible to bound the number of samples required for Hutchinson's method. However, by the bound that we obtain is not as tight as the bound we obtain when the entries of z are i.i.d normal variables.

3. Three and an Half Estimators

In this section we describe the trace estimators that we analyze. We describe three estimators and a variant of one of them. All estimators follow the same basic pattern: a random vector z is drawn from some fixed distribution, and $z^T Az$ is used to estimate the trace. This procedure is repeated M times using i.i.d samples and the estimates are averaged.

The first estimator uses vectors whose entries are standard Gaussian (normal) variables.

Definition 2. A *Gaussian trace estimator* for a symmetric positive-definite matrix $A \in \mathbb{R}^{n \times n}$ is

$$G_M = \frac{1}{M}\sum_{i=1}^{M} z_i^T Az_i \,,$$

where the z_i's are M independent random vectors whose entries are i.i.d standard normal variables.

The Gaussian estimator does not constrain the 2-norm of the z_i's; it can be arbitrarily small or large. All the other estimators that we analyze normalize the quadratic forms by constraining $z^T z$ to be equal to n. This property alone allows us to prove below a general convergence bound.

Definition 3. A *normalized Rayleigh-quotient trace estimator* for a symmetric positive semi-definite matrix $A \in \mathbb{R}^{n\times n}$ is

$$R_M = \frac{1}{M}\sum_{i=1}^{M} z_i^T A z_i \ ,$$

where the z_i's are M independent random vectors such that $z_i^T z_i = n$ and $\mathrm{E}(z_i^T A z_i) = \mathrm{trace}(A)$.

The second estimator we analyze is Hutchinson's.

Definition 4. An *Hutchinson trace estimator* for a symmetric positive-definite matrix $A \in \mathbb{R}^{n\times n}$ is

$$H_M = \frac{1}{M}\sum_{i=1}^{M} z_i^T A z_i \ ,$$

where the z_i's are M independent random vectors whose entries are i.i.d Rademacher random variables.

The first two estimators use a very large sample spaces. The Gaussian estimator uses continuous random variables, and the Hutchinson estimator draws z from a set of 2^n vectors. Thus, the amount of random bits required to form a sample is $\Omega(n)$. Our third estimator samples from a set of n vectors, so it only needs $O(\log n)$ random bits per sample. We discuss the issue of randomness and it implications further in the next section. The third estimator samples from a smaller family by estimating the trace in a more direct way: it samples the diagonal itself. The average value of a diagonal element of A is $\mathrm{trace}(A)/n$. So we can estimate the trace by sampling a diagonal element and multiplying the result by n. This corresponds to sampling a unit vector from the standard basis and computing the Rayleigh quotient.

Definition 5. A *unit vector estimator* for a symmetric positive-definite matrix $A \in \mathbb{R}^{n\times n}$ is

$$U_M = \frac{n}{M}\sum_{i=1}^{M} z_i^T A z_i \ ,$$

where the z_i's are M independent uniform random samples from $\{e_1, \ldots, e_n\}$.

In contrast to previous methods, the quadratic forms in the unit-vector estimator do not depend in any way on the off-diagonal elements of A, only on the diagonal elements. Therefore, the convergence of U_M is independent of the off-diagonal elements. The distribution of diagonal elements does influence, of course, the convergence to $\mathrm{trace}(A)/n$. For some matrices, this method must sample all the diagonal elements for U_M to be close to $\mathrm{trace}(A)$. For example, if A has one huge diagonal element, the average is useless until we sample this particular element. On the other hand, if all the diagonal elements are the same, the average converges to the exact solution after one sample.

Our last estimator is a variant of the unit vector estimator that uses randomization to address this difficulty. Instead of computing the trace of A, it computes the trace of $\mathcal{F}A\mathcal{F}^T$ where $\mathcal{F}$ is a unitary matrix. Since the *mixing matrix* $\mathcal{F}$ is a unitary, $\mathrm{trace}(A) = \mathrm{trace}(\mathcal{F}A\mathcal{F}^T)$. We construct $\mathcal{F}$ using a randomized algorithm that guarantees with high probability a relatively flat distribution of

the diagonal elements of $\mathcal{F}A\mathcal{F}^T$. More precisely, we construct $\mathcal{F}$ in a way that attempts to flatten the distribution of all the elements of $\mathcal{F}A\mathcal{F}^T$, not just its diagonal elements. We use this strategy because we do not know how to flatten the diagonal elements alone. Our constructions are based on the random mixing matrices suggested in [2].

Definition 6. A *random mixing matrix* is a unitary matrix $\mathcal{F} = FD$, where F and D are n-by-n unitary matrices. The matrix F is a fixed unitary matrix called the *seed* matrix. The matrix$D$ is a unitary random diagonal matrix with diagonal entries that are i.i.d Rademacher random variables: $\Pr(D_{ii} = \pm 1) = 1/2$.

Definition 7. A *mixed unit vector estimator* for a symmetric positive semi-definite matrix $A \in \mathbb{R}^{n\times n}$ is

$$T_M = \frac{n}{M}\sum_{i=1}^{M} z_i^T \mathcal{F}A\mathcal{F}^T z_i \,,$$

where the z_i's are M independent uniform random samples from $\{e_1, \ldots, e_n\}$, and $\mathcal{F}$ is a random mixing matrix.

The mixing effectiveness of $\mathcal{F}$ depends on the quantity $\eta = \max |F_{ij}|^2$ [2, 3]. A small η guarantees effective mixing. We discuss this further in section 8.

We choose the fixed seed matrix F so as to minimize $\eta = \max |F_{ij}|^2$. The minimal value of η for a unitary F is $1/n$. A normalized DFT matrix achieves the minimum, but applying it requires complex arithmetic. A normalized Hadamard matrix also achieves the minimum and its entries are real. However, Hadamard matrices do not exist for all dimensions, so they are more difficult to use (they require padding). The Discrete Cosine Transform (DCT) and the Discrete Hartley Transform (DHT), which are real, exist for any dimension, and can be applied quickly, but their η value is $2/n$, twice as large as that of the DFT and the Hadamard. All are valid choices. The decision should be based on the implementation cost of computing columns of F and applying DAD^T to them versus the value of η.

4. Comparing the Quality of Estimators

The easiest way to analyze the quality of trace estimators is to analyze their variance. For any Monte-Carlo estimator R_M we have $\mathrm{Var}(R_M) = \mathrm{Var}(R_1)/M$ so we only need to analyze the variance of a single sample. This type of analysis usually does not reveal much about the estimator, because the variance is usually too large to apply Chebyshev's inequality effectively.

A better way to analyze an estimator is to bound the number of samples required to guarantee that the probability of the relative error exceeding ϵ is at most δ.

Definition 8. Let A be a symmetric positive semi-definite matrix. A randomized trace estimator T is an (ϵ, δ)-*approximator* of $\mathrm{trace}(A)$ if

$$\Pr\left(|T - \mathrm{trace}(A)| \le \epsilon\,\mathrm{trace}(A)\right) \ge 1 - \delta\,.$$

The third metric that we analyze is the number of random bits used by the algorithm, i.e. the randomness of the algorithm. The trace estimators are highly parallel; each Rayleigh quotient can be computed by a separate processor. If the number of random bits is small, they can be precomputed by a sequential random number generator. If the number is large (e.g., $O(n)$ per Rayleigh quotient), the implementation will need to use a parallel random-number generator. This concern is common to all Monte-Carlo methods.

Estimator	Variance of one sample	Bound on # samples for an (ϵ, δ)-approx	Random bits per sample
Gaussian	$2\|A\|_F$	$20\epsilon^{-2}\ln(2/\delta)$	infinite; $\Theta(n)$ in floating point
Normalized Rayleigh-quotient	-	$\frac{1}{2}\epsilon^{-2}n^{-2}\operatorname{rank}^2(A)\ln(2/\delta)\kappa_f^2(A)$	-
Hutchinson's	$2\left(\|A\|_F^2 - \sum_{i=1}^n A_{ii}^2\right)$	$6\epsilon^{-2}\ln(2\operatorname{rank}(A)/\delta)$	$\Theta(n)$
Unit Vector	$n\sum_{i=1}^n A_{ii}^2 - \operatorname{trace}^2(A)$	$\frac{1}{2}\epsilon^{-2}\ln(2/\delta)r_D^2(A)$ $r_D(A) = \frac{n\cdot\max_i A_{ii}}{\operatorname{trace}(A)}$	$\Theta(\log n)$
Mixed Unit Vector	-	$8\epsilon^{-2}\ln\left(4n^2/\delta\right)\ln(4/\delta)$	$\Theta(\log n)$

TABLE 1. Summary of results: quality of the estimators under different metrics. The proofs appear in sections 5-8.

Table 1 summarizes the results of our analyses. The proofs are in sections 5-8. The smallest variance is achieved by Hutchinson's estimator, but the Gaussian estimator has a better (ϵ, δ) bound. Unit vector estimators use the fewest random bits, but have an (ϵ, δ) bound that is worse than that of Gaussian and Hutchinson's estimators.

The (ϵ, δ) bounds are not necessarily tight. Our numerical experiments did not show a considerable difference in practice between the Gaussian, Hutchinson and mixed unit vector estimators. See section 9.

From a theoretical point of view, the (ϵ, δ) bound for the Gaussian estimator seems good; for fixed ϵ and δ, only $O(1)$ samples are needed. However, the ϵ^{-2} factor in the bound implies that the number of samples may need to scale exponentially with the number of bits of accuracy (the number of samples in the bound scales exponentially with $\log_{10}\epsilon^{-1}$). Therefore, in applications that require only a modest ϵ, say $\epsilon = 0.1$, the Gaussian estimator is good. But in applications that require a small ϵ, even $\epsilon = 10^{-3}$, the number of samples required may be too high.

Are these bounds tight? If they are not, the algorithms themselves may be useful even for small ϵ.

Although we do not have a formal lower-bound, we conjecture that our bound on G_M is almost asymptotically tight. Consider the order n all-ones matrix A. This matrix has a single non-zero eigenvalue n and $n-1$ zero eigenvalues. We see that $\frac{1}{n}z^TAz \sim \chi^2(1)$. Therefor $MG_M/n \sim \chi^2(M)$. This means that G_M has mean n and variance $2n^2/M$. The χ^2 distribution is the sum of independent random variables, so by the central limit theorem it converges to a normal distribution for large M. This convergence to normality is rather fast, and $M \geq 50$ degrees of freedom is usually considered sufficient for the χ^2 distribution to be "approximately normal" [7]. We find that

$$\begin{aligned}\Pr(G_M - n \geq \epsilon n) &\approx \operatorname{erfc}(\epsilon\sqrt{M/2}) \\ &\geq \frac{2}{\sqrt{\pi}}\cdot\frac{\exp(-\epsilon^2 M/2)}{\epsilon\sqrt{M/2}+\sqrt{\epsilon^2 M/2+2}}\,,\end{aligned}$$

Let C_δ be the solution to

$$C_\delta\left(\sqrt{\ln(C_\delta/\sqrt{\pi}\delta)}+\sqrt{\ln(C_\delta/\sqrt{\pi}\delta)+2}\right) = 2\,.$$

If $M < 2\epsilon^{-2}\ln(C_\delta/\pi\delta)$ where we find that

$$\begin{aligned}\Pr(G_M - n \geq \epsilon n) &\geq \frac{2}{\sqrt{\pi}} \cdot \frac{\exp(\ln(\sqrt{\pi}\delta/C_\delta))}{\sqrt{\ln(C_\delta/\sqrt{\pi}\delta)} + \sqrt{\ln(C_\delta/\sqrt{\pi}\delta)+2}}\,, \\ &= \frac{2}{C_\delta\left(\sqrt{\ln(C_\delta/\sqrt{\pi}\delta)} + \sqrt{\ln(C_\delta/\sqrt{\pi}\delta)+2}\right)} \cdot \delta \\ &= \delta\,.\end{aligned}$$

The bound is $\Omega(\epsilon^{-2})$ for a fixed δ, but it is not $\Omega(\epsilon^{-2}\ln(1/\delta))$ as $C_\delta \to 0$ if $\delta \to 0$. Nevertheless, this decay is slow and it appears that our bound is almost asymptotically tight.

The main difficulty in turning this argument into a formal proof is the approximation phase $\Pr(G_M - n \geq \epsilon n) \approx \operatorname{erfc}(\epsilon\sqrt{M/2})$. While it is true that the χ^2 distribution converges to the normal distribution, convergence can be very slow. Indeed, the Berry-Esseen Theorem [8, § 16.5] guarantees a convergence rate that is proportional only to $M^{-1/2}$. So for a fixed δ there exists an ϵ that is small enough such that the sample size will be so large that the tail bound on normal approximation kicks in. Indeed every Monte-Carlo i.i.d estimator with non-zero finite variance converges to a normal distribution, but the general wisdom on the χ^2 distribution is that it converges very quickly to the normal distribution.

A more direct way to prove a lower bound will be to use some lower bound on the tail of the chi-squared cumulative distribution function. Unfortunately, current bounds ([15, 11]) are too complex to provide a useful lower bound, and deriving a simple lower bound is outside the scope of this paper.

In the next section we present experiments that show that convergence rate (in terms of digits of accuracy) on the all-ones matrix is indeed slow, supporting our conjecture that our bound is almost tight.

5. Analysis of the Gaussian Estimator

In this section we analyze the Gaussian estimator. We begin with the variance.

Lemma 9. *Let A be an $n \times n$ symmetric matrix. The single sample Gaussian estimator G_1 of A is an unbiased estimator of* $\operatorname{trace}(A)$ *i.e.,* $\mathrm{E}(G_1) = \operatorname{trace}(A)$ *and* $\operatorname{Var}(G_1) = 2\|A\|_F^2$.

Proof. A is symmetric so it can be diagonalized. Let $\Lambda = UAU^T$ be the unitary diagonalization of A (its eigendecomposition), and define $y = Uz$, where $G_1 = z^T Az$. We can write $G_1 = \sum_{i=1}^n \lambda_i y_i^2$ where y_i is the ith entry of y_i. Since U is unitary, the entries of y are i.i.d Gaussian variables, like the entries of z, so $\mathrm{E}(y_i^2) = 1$ and $\operatorname{Var}(y_i^2) = 2$. We find that

$$\begin{aligned}\mathrm{E}(G_1) &= \sum_{i=1}^n \lambda_i \,\mathrm{E}(y_i^2) = \sum_{i=1}^n \lambda_i = \operatorname{trace}(A)\,, \\ \operatorname{Var}(G_1) &= \sum_{i=1}^n \lambda_i^2 \operatorname{Var}(y_i^2) = 2\sum_{i=1}^n \lambda_i^2 = 2\|A\|_F^2\,.\end{aligned}$$

□

Next, we prove an (ϵ, δ) bound for the Gaussian estimator.

Theorem 10. *Let A be an $n \times n$ symmetric semidefinite matrix. The Gaussian estimator G_M is an (ϵ, δ)-approximator of* $\operatorname{trace}(A)$ *for $M \geq 20\epsilon^{-2}\ln(2/\delta)$.*

Proof. A is symmetric so it can be diagonalized. Let $\Lambda = UAU^T$ be the unitary diagonalization of A (its eigendecomposition), and define $y_i = Uz_i$. Since U is unitary, the entries of y_i are i.i.d Gaussian variables. Notice that $G_M = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{n}\lambda_j y_{ij}^2 = \frac{1}{M}\sum_{j=1}^{n}\lambda_j\sum_{i=1}^{M}y_{ij}^2$ where y_{ij} is the jth entry of y_i.

We prove the bound using a Chernoff-style argument. y_{ij} is a standard normal random variable so $\sum_{i=1}^{M}y_{ij}^2$ is χ^2 with M degrees of freedom. Therefore, the moment generating function of $Z = MG_M$ is

$$\begin{aligned} m_Z(t) &= \mathrm{E}(\exp(tZ)) \\ &= \prod_{i=1}^{n}(1-2\lambda_i t)^{-M/2} \\ &= (1-2\tau t + h(t))^{-M/2} \end{aligned} \tag{5.1}$$

where

$$\tau = \mathrm{trace}(A)$$

and

$$h(t) = \sum_{s=2}^{n}(-2)^s t^s \sum_{\substack{S\subseteq\Lambda \\ |S|=s}}\prod_{x\in S}x$$

as long as $|\lambda_i t| \le \frac{1}{2}$ for all i (Λ is the set of A's eigenvalues).

It is easy to see if $\{x_1,\ldots,x_n\}$ is a set of non-negative real numbers, then for all $i = 1,\ldots,n$ we have

$$\sum_{\substack{S\subseteq[n] \\ |S|=i}}\prod_{j\in S}x_j \le \left(\sum_{i=1}^{n}x_i\right)^i,$$

where $[n] = \{1,\ldots,n\}$. Therefore, we can bound

$$|h(t)| \le \sum_{j=2}^{n}(2\tau t)^j\,.$$

Set $t_0 = \epsilon/(4\tau(1+\epsilon/2))$. For all i we have $\lambda_i t_0 \le \frac{1}{2}$, so (5.1) is the correct formula for $m_Z(t_0)$. We now have

$$\begin{aligned} |h(t_0)| &\le \sum_{j=2}^{n}\left(\frac{\epsilon}{2(1+\epsilon/2)}\right)^j \\ &\le \frac{\epsilon^2}{4(1+\epsilon/2)^2}\cdot\frac{1}{1-\frac{\epsilon}{2(1+\epsilon/2)}}. \\ &= \frac{\epsilon^2}{4(1+\epsilon/2)} \end{aligned}$$

Markov's inequality asserts that

$$\begin{aligned}
\Pr\left(G_M \geq \tau(1+\epsilon)\right) &= \Pr\left(Z \geq \tau M(1+\epsilon)\right). \\
&\leq m_Z(t_0)\exp(-\tau M(1+\epsilon)t_0) \\
&\leq \left(1-\epsilon/2\,(1+\epsilon/2)-\epsilon^2/4(1+\epsilon/2)\right)^{-M/2}\cdot\exp(-\frac{M}{2}\cdot\frac{\epsilon}{2}\cdot\frac{1+\epsilon}{1+\epsilon/2}) \\
&= \exp(-\frac{M}{2}(\ln(1-\epsilon/2(1+\epsilon/2)-\epsilon^2/4(1+\epsilon/2))+\frac{\epsilon}{2}\cdot\frac{1+\epsilon}{1+\epsilon/2})) \\
&= \exp(-\frac{M}{2}(\ln(1-\epsilon/2)+\frac{\epsilon}{2}\cdot\frac{1+\epsilon}{1+\epsilon/2})) \\
&= \exp\left(-\frac{M}{2}\left(\frac{\epsilon}{2}\cdot\frac{1+\epsilon}{1+\epsilon/2}-\sum_{i=1}^{\infty}\frac{(\epsilon/2)^i}{i})\right)\right) \\
&= \exp\left(-\frac{M}{2}\left(\frac{\epsilon}{2}\left(\frac{1+\epsilon}{1+\epsilon/2}-1\right)-\frac{\epsilon^2}{8}-\frac{\epsilon^2}{4}\sum_{i=1}^{\infty}\frac{(\epsilon/2)^i}{(i+2)})\right)\right) \\
&\leq \exp\left(-\frac{M}{2}\left(\frac{\epsilon^2}{4}\cdot\frac{1}{1+\epsilon/2}-\frac{\epsilon^2}{8}+\frac{\epsilon^2}{4}\ln(1-\epsilon/2)\right)\right) \\
&= \exp\left(-\frac{M\epsilon^2}{8}\left(\frac{1}{1+\epsilon/2}-\frac{1}{2}+\ln(1-\epsilon/2)\right)\right) \\
&\leq \exp(-M\epsilon^2/20)
\end{aligned}$$

for $\epsilon \leq 0.1$. We find that if $M \geq 20\epsilon^{-2}\ln(2/\delta)$ then $\Pr\left(G_M \leq \tau(1+\epsilon)\right) \leq \delta/2$. Using the same technique a lower bound can be shown, and combined with a union-bound we find that $\Pr\left(|G_M - \tau| \leq \tau(1+\epsilon)\right) \leq \delta$. □

In some cases it is possible to prove better bounds, or even the exact trace. For example, we show that using a Gaussian trace estimator we can compute the rank of a projection matrix (i.e., a matrix with only 0 and 1 eigenvalues) using only $O(\mathrm{rank}(A)\log(2/\delta))$ samples (where δ is a probability of failure; there is no dependence on ϵ). Finding the rank of a projection matrix is useful for computing charge densities (in electronic structures calculations) without diagonalization [6].

Lemma 11. *Let $A \in \mathbb{R}^{n\times n}$ be a projection matrix, and let $\delta > 0$ be a failure probability. For $M \geq 24\,\mathrm{rank}(A)\ln(2/\delta)$, the Gaussian trace estimator G_M of A satisfies*

$$\Pr(\mathrm{round}(G_M) \neq \mathrm{rank}(A)) \leq \delta\,.$$

Proof. A projection matrix has only 0 and 1 eigenvalue, so the eigenvalue decomposition of A is of the form

$$A = U^T \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} U\,.$$

If we write $y = Uz$ then $z^T Az = \sum_{i=1}^{\operatorname{rank}(A)} y_i^2$. Since U is unitary the entries of y_i are i.i.d Gaussian variables, so $z^T Az$ is χ^2 with $\operatorname{rank}(A)$ degrees of freedom. The χ^2 distribution is additive, so MG_M is also χ^2 but with $M \operatorname{rank}(A)$ degrees of freedom. We now use a known tail-bounds on the χ^2 distribution [12]: if $X \sim \chi^2(k)$ then

$$\Pr(|X - k| \leq \epsilon k) \leq 2 \exp(-k\epsilon^2/6) .$$

By applying this result to MG_M we find that

$$\begin{aligned} \Pr(|G_M - \operatorname{rank}(A)| \geq \operatorname{rank}(A)\epsilon) &= \Pr(|MG_M - M \operatorname{rank}(A)| \geq M \operatorname{rank}(A)\epsilon) \\ &\leq 2 \exp(-M \operatorname{rank}(A)\epsilon^2/6) . \end{aligned}$$

If we set

$$M \geq 6 \operatorname{rank}(A)^{-1} \epsilon^{-2} \ln(2/\delta) \tag{5.2}$$

we find that

$$\Pr(|G_M - \operatorname{rank}(A)| \geq \operatorname{rank}(A)\epsilon) \leq \delta .$$

If A is a projection matrix, then $\operatorname{trace}(A) = \operatorname{rank}(A)$ is an integer, so if the error is below $\frac{1}{2}$, then $\operatorname{round}(G_M) = \operatorname{rank}(A)$. We set $\epsilon = 1/(2 \operatorname{rank}(A))$ and obtain

$$\Pr(\operatorname{round}(G_M) \neq \operatorname{rank}(A)) = \Pr(|G_M - \operatorname{rank}(A)| \geq \operatorname{rank}(A)\epsilon) \leq \delta .$$

If we plug ϵ into (5.2) we find that we require $M \geq 24 \operatorname{rank}(A) \ln(2/\delta)$. □

6. General Bound for Normalized Rayleigh quotient Estimators

The sample vectors z in the Gaussian estimator are not normalized, and this can lead to a large $z^T Az$ (but only with a small probability). Normalized estimators are somewhat easier to analyze because each sample is bounded. When A is well conditioned, we get a useful and very general bound.

Theorem 12. *A normalized Rayleigh estimator R_M is an (ϵ, δ)-approximator of* $\operatorname{trace}(A)$ *for $M \geq \frac{1}{2}\epsilon^{-2} n^{-2} \operatorname{rank}^2(A) \ln(2/\delta) \kappa_f^2(A)$, where $\kappa_f(A)$ is the ratio between the largest and smallest nonzero eigenvalue of A.*

Proof. Let $0 = \lambda_1 = \cdots = \lambda_k \leq \cdots \leq \lambda_n$ be the eigenvalues of A where $k = n - \operatorname{rank}(A) + 1$, so $\kappa_f(A) = \lambda_n/\lambda_k$. It is easy to see that

$$\begin{aligned} \operatorname{trace}(A) \cdot \kappa_f(A) &= \sum_{i=1}^{n} \lambda_i \cdot \kappa_f(A) \\ &= \sum_{i=k}^{n} \frac{\lambda_i}{\lambda_k} \lambda_n \\ &\geq (n - k + 1)\lambda_n \\ &= \operatorname{rank}(A)\lambda_n \end{aligned}$$

therefore for all i

$$0 \leq z_i^T A z_i \leq \lambda_n z_i^T z_i = n\lambda_n \leq \frac{n}{\operatorname{rank}(A)} \operatorname{trace}(A) \cdot \kappa_f(A) .$$

According to Hoeffding's inequality for any $t > 0$,

$$\Pr(|R_M - \operatorname{trace}(A)| \geq t) \leq 2\exp\left(-\frac{2M^2 \operatorname{rank}^2(A) t^2}{M \cdot n^2 \operatorname{trace}^2(A)\kappa_f^2(A)}\right).$$

If we set $t = \epsilon \operatorname{trace}(A)$ we find that

$$\Pr(|R_M - \operatorname{trace}(A)| \geq \epsilon \operatorname{trace}(A)) \leq 2\exp\left(-\frac{2M \operatorname{rank}^2(A)\epsilon^2}{n^2\kappa_f^2(A)}\right).$$

We now set M so that the bound is smaller than δ:

$$\frac{2M \operatorname{rank}^2(A)\epsilon^2}{n^2\kappa_f^2(A)} \geq \ln\left(\frac{2}{\delta}\right)$$

or

$$M \geq \frac{\ln(2/\delta) \cdot n^2\kappa_f^2(A)}{2\operatorname{rank}^2(A)\epsilon^2}.$$

□

7. Analysis of Hutchinson's Estimator

When A is ill conditioned, the (ϵ, δ) bound in Section 6 is weak. We can sharpen it for a specific normalized estimator, that of Hutchinson. However, the bound is still weaker than that of the Gaussian estimator. The bound here is of interest because (1) Hutchinson's estimator is widely used, (2) it uses fewer random bits than the Gaussian estimator, and (3) it requires only additions and subtractions, not multiplications. It is also possible that there is an even stronger bound for Hutchinson's method.

Theorem 13. *The Hutchinson estimator H_M is an (ϵ, δ)-approximator of* $\operatorname{trace}(A)$ *for* $M \geq 6\epsilon^{-2}\ln(2\operatorname{rank}(A)/\delta)$.

To prove this theorem we use the following Lemma from [1, Lemma 5]:

Lemma 14. *Let $\alpha \in \mathbb{R}^n$ be an arbitrary unit vector. Define $Q = (\alpha^T z)^2$ where z is a random vector whose entries are i.i.d Rademacher random variables ($\Pr(z_i = \pm 1) = 1/2$). Let $Q_1, \ldots, Q_M$ be M i.i.d copies of Q (different zs but the same α), and define $S = \frac{1}{M}\sum_{i=1}^M Q_i$. Then, for any $\epsilon > 0$,*

$$\Pr(|S - 1| \geq \epsilon) \leq 2\exp\left(-\frac{M}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)\right).$$

Proof. (of Theorem 13). A is symmetric and semidefinite so it can be diagonalized. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A and assume without loss of generality that the non-zero eigenvalues are $\lambda_1, \ldots, \lambda_{\operatorname{rank}(A)}$. Let $\Lambda = UAU^T$ be the unitary diagonalization of A (its eigendecomposition), and define $y_i = U^T z_i$. Notice that $H_M = \frac{1}{M}\sum_{i=1}^M \sum_{j=1}^n \lambda_j y_{ij}^2 = \sum_{j=1}^n \lambda_j \frac{1}{M}\sum_{i=1}^M y_{ij}^2$ where y_{ij} is the jth entry of y_i. The rows U_i^T of U^T are unit vectors so $S = \frac{1}{M}\sum_{i=1}^M \left(U_j^T z_i\right)^2$ satisfies the conditions of Lemma 14. But we also have $S = \frac{1}{M}\sum_{i=1}^M y_{ij}^2$, so

$$\Pr\left(\left|\frac{1}{M}\sum_{i=1}^M y_{ij}^2 - 1\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{M}{2}\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)\right).$$

If $M \geq 6\epsilon^{-2}\ln(2\operatorname{rank}(A)/\delta)$ this implies that

$$\Pr\left(\left|\frac{1}{M}\sum_{i=1}^{M} y_{ij}^2 - 1\right| \geq \epsilon\right) \leq \frac{\delta}{\operatorname{rank}(A)}\,.$$

This bound holds for each specific j. Using the union-bound, we conclude that the probability that the error is larger than ϵ for some $j = 1, \ldots, \operatorname{rank}(A)$ is at most δ. Hence, the probability that the error is smaller than ϵ for all $j = 1, \ldots, \operatorname{rank}(A)$ is at least $1 - \delta$. So with probability $1 - \delta$ we also have

$$\begin{aligned}
|H_M - \operatorname{trace}(A)| &= \left|\sum_{j=1}^{n} \lambda_j \frac{1}{M}\sum_{i=1}^{M} y_{ij}^2 - \sum_{i=1}^{n}\lambda_i\right| \\
&= \left|\sum_{j=1}^{\operatorname{rank}(A)} \lambda_j \left(\frac{1}{M}\sum_{i=1}^{M} y_{ij}^2 - 1\right)\right| \\
&\leq \sum_{j=1}^{\operatorname{rank}(A)} \lambda_j \left|\frac{1}{M}\sum_{i=1}^{M} y_{ij}^2 - 1\right| \\
&\leq \epsilon \sum_{j=1}^{\operatorname{rank}(A)} \lambda_j \\
&= \epsilon \operatorname{trace}(A)\,.
\end{aligned}$$

□

The bound is larger than the bound for the Gaussian estimator by a $\ln(\operatorname{rank}(A))$ factor. The main difficulty here is that, unlike the Gaussian estimator, the Hutchinson's estimator cannot be written as a weighted sum of i.i.d random variables. This forces us to use a union bound instead of using a global analysis. Nevertheless, given the better variance term of Hutchinson's estimator we conjecture that this $\ln(\operatorname{rank}(A))$ factor is redundant. In fact, there are some matrix classes for which Hutchinson's estimator is clearly better than the Gaussian estimator. For example, on diagonal or nearly diagonal matrices the Hutchinson's estimator will converge very fast, which is not true for the Gaussian estimator. Another interesting example is the all-ones matrix for which the bound for the Hutchinson estimator is the same as the bound for the Gaussian estimator (it is possible to show that for the all-ones matrix the Gaussian estimator is an (ϵ, δ)-approximator for $M \geq 6\epsilon^{-2}\ln(2/\delta)$).

8. Reducing Randomness: Analyzing Unit Vector Estimators

This section analyzes two unit vector estimators: the unit vector estimator and the mixed unit vector estimator. These estimators' main advantage is in restricting the sample space to n vectors. Thus, only $\lceil \log_2 n \rceil$ random bits are required per sample. This allows the samples to be generated in advance. We begin by analyzing the variance.

Lemma 15. *Let A be an $n \times n$ symmetric matrix. The single sample unit vector estimator U_1 of A is an unbiased estimator of* $\operatorname{trace}(A)$ *i.e.,* $\mathrm{E}(U_1) = \operatorname{trace}(A)$ *and* $\operatorname{Var}(U_1) = n\sum_{i=1}^{n} A_{ii}^2 - \operatorname{trace}^2(A)$ *..*

Proof. Let $U_1 = nz^T Az$. Because z is an identity vector $z^T Az$ just samples values from the diagonal. Every diagonal value is sampled with equal probability, so $\mathrm{E}(z^T Az) = \mathrm{trace}(A)/n$, from which $\mathrm{E}(nz^T Az) = \mathrm{trace}(A)$ follows immediately.

As for variance the following equality holds

$$\begin{aligned}\mathrm{Var}(nz^T Az) &= \mathrm{E}((nz^T Az)^2) - (\mathrm{E}(nz^T Az))^2 \\ &= n^2\,\mathrm{E}((z^T Az)^2) - \mathrm{trace}^2(A)\end{aligned}$$

The random variable $(z^T Az)^2$ samples the square of the diagonal values of A so $\mathrm{E}((z^T Az)^2) = \sum_{i=1}^n A_{ii}^2/n$ and the equality follows. □

We now turn to the more interesting analysis of the number of samples that guarantee an (ϵ, δ)-approximator. This quantity depends on the ratio between the largest possible estimate (when estimating the maximal diagonal value) and the trace.

Theorem 16. *The unit vector estimator U_M is an (ϵ, δ)-approximator of* $\mathrm{trace}(A)$ *for* $M \geq \frac{1}{2}\epsilon^{-2}\ln(2/\delta) r_D^2(A)$ *where* $r_D(A) = \frac{n \cdot \max_i A_{ii}}{\mathrm{trace}(A)}$.

Proof. The unit vector estimator samples values from the diagonal and multiplies them by n, so a single samples takes values in the range $[0, n \cdot \max_i A_{ii}]$. According to Hoeffding's inequality

$$\Pr(|U_M - \mathrm{trace}(A)| \geq t) \leq 2\exp\left(-\frac{2M^2t^2}{Mn^2 \cdot (\max_i A_{ii})^2}\right).$$

If we set $t = \epsilon\,\mathrm{trace}(A)$ we find that

$$\Pr(|U_M - \mathrm{trace}(A)| \geq \epsilon\,\mathrm{trace}(A)) \leq 2\exp\left(-\frac{2M\epsilon^2}{r_D^2(A)}\right).$$

We now set M so that the bound is smaller than δ:

$$\frac{2M\epsilon^2}{r_D^2(A)} \geq \ln\left(\frac{2}{\delta}\right)$$

or

$$M \geq \frac{\ln(2/\delta) \cdot r_D^2(A)}{2\epsilon^2}.$$

□

We now analyze the mixed unit vector estimator. The unit vector estimator relies on the the mixing matrix $\mathcal{F}$. The analysis is based on a lemma from [2, 3].

Lemma 17. *Let U be an $n \times m$ matrix with orthonormal columns, and let $\mathcal{F} = FD$ be a random mixing matrix. With probability of at least $1 - \delta$ ($\delta > 0$) we have for all i and j*

$$\left|(\mathcal{F}U)_{ij}\right| \leq \sqrt{2\eta\ln\left(\frac{2mn}{\delta}\right)},$$

where $\eta = \max|F_{ij}|^2$.

The mixing matrix prevents entries from an orthonormal matrix to be too large. When applied from both sides to a symmetric positive semidefinite matrix it prevents the diagonal elements from being too big, i.e. $r_D(\mathcal{F}A\mathcal{F}^T)$ is not too big.

Theorem 18. *The mixed unit vector estimator T_M is an (ϵ, δ)-approximator of* $\operatorname{trace}(A)$ *for* $M \geq 2n^2\eta^2\epsilon^{-2}\ln(4/\delta)\ln^2(4n^2/\delta)$.

Proof. A is symmetric so it can be diagonalized. Let $\Lambda = U^T AU$ be the unitary diagonalization of A (its eigendecomposition), and let $V = \mathcal{F}U$. It is easy to see that

$$\left(\mathcal{F}A\mathcal{F}^T\right)_{jj} = \sum_{k=1}^{n} \lambda_k V_{jk}^2 \,.$$

According to Lemma 17, with probability $1 - \delta/2$ we have

$$V_{jk}^2 = \left|\left(\mathcal{F}U\right)_{jk}\right|^2 \leq 2\eta \ln\left(\frac{2n^2}{\delta/2}\right) = 2\eta \ln\left(\frac{4n^2}{\delta}\right) \,. \tag{8.1}$$

The eigenvalues λ_i are non-negative, so we conclude that with probability $1 - \delta/2$ for all j,

$$\begin{aligned} 0 \leq \left(\mathcal{F}A\mathcal{F}^T\right)_{jj} &\leq 2\eta \ln\left(\frac{4n^2}{\delta}\right) \sum_{j=1}^{n} \lambda_j \\ &= 2\eta \ln\left(\frac{4n^2}{\delta}\right) \operatorname{trace}(A) \,. \end{aligned}$$

We find that

$$r_D(\mathcal{F}A\mathcal{F}^T) \leq 2n\eta \ln\left(\frac{4n^2}{\delta}\right) \,.$$

Therefore, according to Theorem 16, for $M \geq 2n^2\eta^2\epsilon^{-2}\ln(4/\delta)\ln^2(4n^2/\delta)$ we have

$$\Pr\left(|T_M - \operatorname{trace}(A)| > \epsilon \operatorname{trace}(A)\right) \leq 1 - \delta/2 \,.$$

There can be failures of two kinds: with probability at most $\delta/2$ the bound on the diagonal elements of the mixed matrix may fail to hold, and even if it holds, with probability $\delta/2$ the ϵ bound on the estimation error may fail to hold. We conclude that with probability $1 - \delta$ the error bound does hold. □

Remark 19. For Fourier-type matrices, such as DFT and DCT, $\eta = \Theta(1/n)$, so the lower bound on M becomes simpler,

$$M \geq C \frac{\ln^2\left(4n^2/\delta\right)\ln(4/\delta)}{\epsilon^2} \,,$$

for some small C (8 for the case of DCT, 2 for DFT).

9. Experiments

We present the results of several computational experiments that compare the different estimators, and clarify the actual convergence rate.

Figure 9.1 shows the convergence of the various estimators on a matrix of order $n = 100,000$ whose elements are all 1. We have used this matrix as an example of the matrix with the largest variance possible for Hutchinson's and Gaussian estimator. The graphs show that all methods converge quite slowly. There is no significant difference in the convergence behavior of all three methods, although we presented different bounds. The graph also supports our conjecture that our bounds are almost tight, and that the cost is exponential in the number of required accuracy digits.

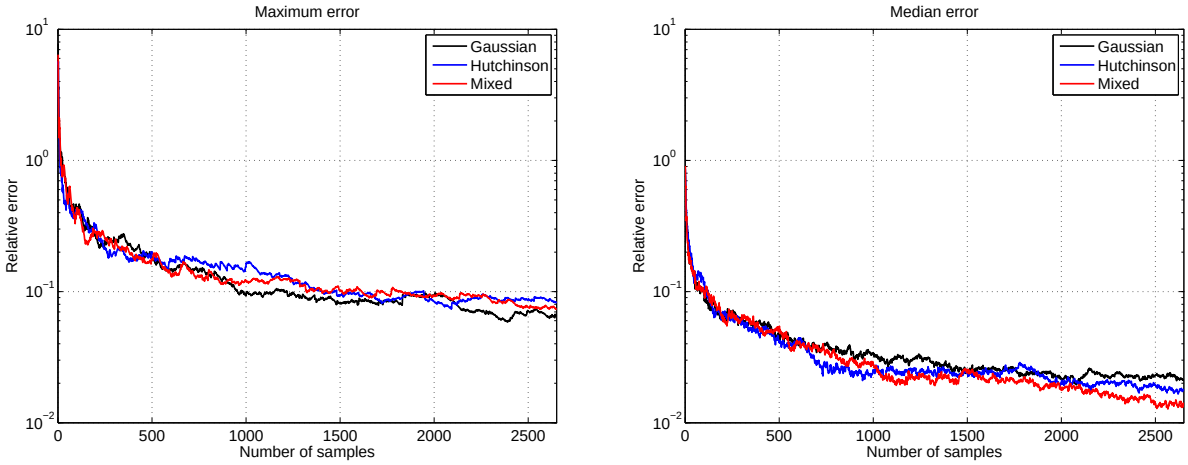

FIGURE 9.1. Convergence of the estimators on a matrix of order $100,000$ whose elements are all 1. The graph on the left shows the maximum error during 100 runs of the algorithm, and the graph on the right the median of the 100 runs.

Figure 9.2 clarifies the convergence behavior of the estimators. The graph on the left shows the convergence all the way up to n iterations, with two variants of the mixed estimator: with and without repetitions. Convergence stagnates and the error nears machine ϵ only very close to iteration n and only when sampling without repetitions. If we sample without repetitions, after we sample all the sample space we are guaranteed to have the exact trace (this is not possible for the Gaussian estimator and Hutchinson's estimator, but also not practical in our method). The histogram on the right show that in spite of the mixing that $\mathcal{F}$ performs, the diagonal elements of the mixed matrix $\mathcal{F}A\mathcal{F}^T$ are still highly skewed. In other words, there are some diagonal values that are important to sample; until they are sampled, the error remains large.

Figure 9.3 shows that on other classes of matrices, the methods reach a smaller error before they stagnate. On a random dense matrix, the methods converge quickly to an error smaller than 10^{-2}, but then stagnate. On a sparse matrix from the University of Florida matrix collection, the methods reach an error of about 10^{-3} and then stagnate. There is again little difference between the convergence rates of the three methods, although it seems that Gaussian estimator is a little less accurate then the other two estimators.

10. Conclusions

In terms of the (ϵ, δ) bounds, the Gaussian estimator, requires the smallest number of samples. The convergence bound for Hutchinson's estimator is the runner up: it requires more iterations than the Gaussian, but fewer than the mixed unit vector estimator.

In terms of the number of random bits that these estimators require, the ranking is the exact opposite: the Gaussian estimator requires the most bits, followed by Hutchison's estimator, and the mixed unit vector estimator requires the least.

Convergence to a small error is slow, both in practice and in terms of the bounds. The ϵ^{-2} factor in all the bounds imply that the number of samples required to get close to, say, machine epsilon, is huge. The estimators quickly give a crude estimate of the trace (correct to within 0.1 or 0.01, say), but they require a huge number of samples to obtain a very accurate estimate.

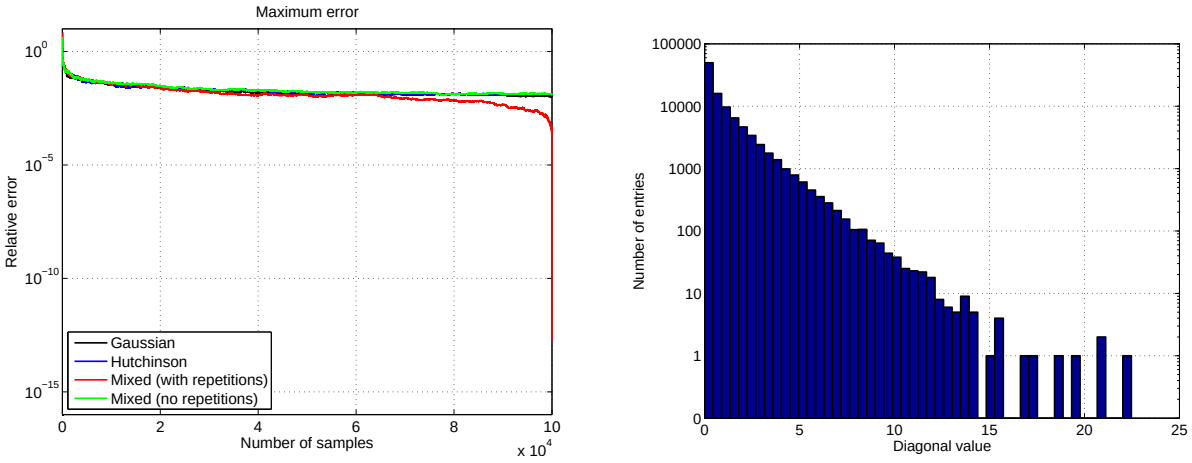

FIGURE 9.2. Details to clarify the behavior of the methods. The experiment is similar to the one in Figure 9.1. The graph on the left shows convergence all the way to n iterations, and the histogram on the right shows the distribution of diagonal values (relevant for the estimator presented in section 8).

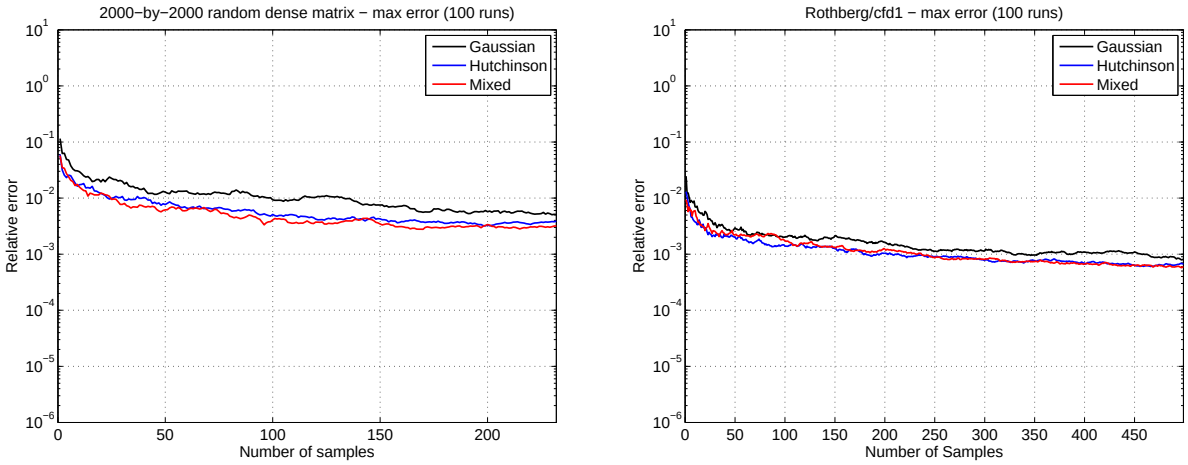

FIGURE 9.3. Convergence on two more matrices: a random matrix of order 2000 (left) and a sparse matrix of order $70,656$.

The ϵ^{-2} factor in the bound is common to many Monte-Carlo algorithms in numerical linear algebra. When the Monte-Carlo method is used as an inexact solver within the context of an iterative solver, the overall algorithm can be both fast and accurate [3]. We are not aware of a suitable iterative algorithm for trace computations.

ACKNOWLEDGMENTS

Acknowledgement. It is a pleasure to thank Mark Tygert for helpful comments and ideas.

References

[1] Dimitris Achlioptas. Database-friendly random projections. In *PODS '01: Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database systems*, pages 274–281, New York, NY, USA, 2001. ACM.

[2] Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *STOC '06: Proceedings of the thirty-eighth annual ACM Symposium on Theory of Computing*, pages 557–563, New York, NY, USA, 2006. ACM.

[3] Haim Avron, Petar Maymounkov, and Sivan Toledo. Blendenpik: Supercharging lapack's least-squares solver. *SIAM Journal on Scientific Computing*, 32(3):1217–1236, 2010.

[4] Z. Bai, M. Fahey, G. Golub, M. Menon, and E. Richter. Computing partial eigenvalue sum in electronic structure calculations. Technical Report SCCM-98-03, Stanford University, Jan 1998.

[5] Zhaojun Bai, Mark Fahey, and Gene Golub. Some large scale matrix computation problems. *J. Comput. Appl. Math*, 74:71–89, 1996.

[6] C. Bekas, E. Kokiopoulou, and Y. Saad. An estimator for the diagonal of a matrix. *Appl. Numer. Math.*, 57(11-12):1214–1229, 2007.

[7] George E. P. Box, William G. Hunter, and J. Stuart Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. John Wiley & Sons, June 1978.

[8] William Feller. *An Introduction to Probability Theory and Its Applications, Vol. 2*. Wiley, 3rd. edition, January 1971.

[9] M. F. Hutchinson. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics, Simulation and Computation*, (18):1059–1076, 1989.

[10] Toshiaki Iitaka and Toshikazu Ebisuzaki. Random phase vector for calculating the trace of a large matrix. *Physical Review E*, 69:057701–1–057701–4, 2004.

[11] A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and B. Zwart. Gaussian expansions and bounds for the Poisson distribution applied to the Erlang B formula. *Advances in Applied Probability*, 40(1):122–143, 2008.

[12] Ping Li, Trevor Hastie, and Kenneth Church. Nonlinear estimators and tail bounds for dimension reduction in l_1 using Cauchy random projections. In Nader H. Bshouty and Claudio Gentile, editors, *Learning Theory*, volume 4539 of *Lecture Notes in Computer Science*, chapter 37, pages 514–529. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[13] R. N. Silver and H. Röder. Calculation of densities of states and spectral functions by chebychev recursion and maximum entropy. *Physical Review E*, 56:4822–4829, 1997.

[14] Charalampos E. Tsourakakis. Fast counting of triangles in large real networks without counting: Algorithms and laws. *IEEE International Conference on Data Mining (ICDM 2008)*, 0:608–617, 2008.

[15] David L. Wallace. Bounds on normal approximations to student's and the chi-square distributions. *The Annals of Mathematical Statistics*, 30(4):1121–1130, 1959.

[16] Mei Ning Wong, F. J. Hickernell, and Kwong Ip Liu. Computing the trace of a function of a sparse matrix via Hadamard-like sampling. 2004.