Clifford algebras and approximating the permanent[†]

Steve Chien[‡] Lars Rasmussen[§] Alistair Sinclair[¶]

July 29, 2005

Abstract

We study approximation algorithms for the permanent of an $n \times n$ $(0, 1)$ matrix A based on the following simple idea: obtain a random matrix B by replacing each 1-entry of A independently by $\pm e$, where e is a random basis element of a suitable algebra; then output $|\det(B)|^2$. This estimator is always unbiased, but it may have exponentially large variance. In our first main result we show that, if we take the algebra to be a *Clifford algebra* of dimension polynomial in n, then we get an estimator with small variance. Hence only a constant number of trials suffices to estimate the permanent to good accuracy. The idea of using Clifford algebras is a natural extension of earlier work by Godsil and Gutman, Karmarkar *et al.*, and Barvinok, who used the real numbers, complex numbers and quaternions respectively.

The above result implies that, in principle, this approach gives a *fully-polynomial randomized approximation scheme* for the permanent, provided $|\det(B)|^2$ can be efficiently computed in the Clifford algebras. Since these algebras are non-commutative it is not clear how to do this. However, our second main result shows how to compute in polynomial time an estimator with the same mean and variance over the 4-dimensional algebra (which is the quaternions, and is non-commutative); in addition to providing some hope that the computations can be performed in higher dimensions, this quaternion algorithm provides an exponential improvement in the variance over that of the 2-dimensional complex version studied by Karmarkar *et al.*

[†]A preliminary version of this paper appeared in *Proceedings of the 34th Annual Symposium on Theory of Computing* (STOC), 2002, pp. 222–231.

[‡]Computer Science Division, University of California, Berkeley CA 94720-1776, U.S.A. Email: schien@cs.berkeley.edu. Supported by an NSF Graduate Fellowship and NSF grants CCR-9820951 and CCR-0121555.

[§]Digital Fountain, Inc., 39141 Civic Center Drive, Ste. 300, Fremont, CA 94538, U.S.A. Email: lars@digitalfountain.com. Work done while at Computer Science Division, UC Berkeley. Supported by NSF grant CCR-9505448.

[¶]Computer Science Division, University of California, Berkeley CA 94720-1776, U.S.A. Email: sinclair@cs.berkeley.edu. Supported in part by NSF grants CCR-9820951 and CCR-0121555 and DARPA cooperative agreement F30602-00-2-0601.

1 Introduction

The permanent of an $n \times n$ $(0,1)$ matrix $A = (a_{ij})$ is defined as

$$\mathrm{per}(A) = \sum_{\pi \in S_n} \prod_{i=1}^{n} a_{i,\pi i}.$$

Equivalently, per(A) counts the perfect matchings in the $(n+n)$-vertex bipartite graph whose adjacency matrix is A. Computing per(A) exactly is #P-complete, as was shown in Valiant's seminal 1979 paper [20].

The past decade or so has seen a surprising variety of approaches aimed at designing a polynomial time approximation algorithm for the permanent. These can be divided into (at least) four categories: elementary recursive algorithms [18]; reductions to determinants [6, 12, 7, 3, 4]; iterative balancing [15]; and Markov chain Monte Carlo [5, 9, 11, 13, 10]. All the approaches have yielded non-trivial results (at a minimum, fully polynomial approximation schemes for random matrices, or polynomial time approximation algorithms with approximation ratio c^n for a modest constant $c > 1$), and fascinating insights both into the problem and into the wider implications of the associated mathematical techniques. Recently, as reported in [10], the Markov chain Monte Carlo approach led to the first *fully polynomial randomized approximation scheme* for the permanent of an arbitrary $(0,1)$ matrix (and indeed of any matrix with non-negative entries). This is a randomized algorithm which takes as inputs A and a parameter $\varepsilon \in (0,1]$ and outputs a value that approximates per(A) within a factor $1 \pm \varepsilon$ with high probability; the running time is polynomial in n and $\frac{1}{\varepsilon}$.

In this paper we pursue another of the approaches mentioned above, namely reduction to determinants. We are motivated both by the intrinsic elegance of this approach, and by the fact that, if successful, it seems likely to lead to a more efficient algorithm. (The authors of [10] did not attempt to minimize the exponent in their polynomial running time; but even with fine tuning that algorithm is unlikely to be practical.)

The origins of the determinant approach go back to the following beautiful observation of Godsil and Gutman [6]. Let A be an $n \times n$ $(0,1)$ matrix, and let B be the matrix obtained by replacing each 1-entry of A independently by a uniform random element of $\{\pm 1\}$. Then the random variable $(\det(B))^2$ is an *unbiased estimator* of per(A), i.e., its expectation is exactly per(A). This is easy to verify using the facts that the terms in the expansions of permanent and determinant are identical up to sign, and that every cross term in the expansion of $\det(B)^2$ disappears because it contains an independent factor b_{ij} with $\mathrm{E}[b_{ij}] = 0$.

Unfortunately, however, the above estimator has in general a very large variance, so if we were to use it to estimate per(A) we would need to take the mean of exponentially many independent samples to get a good estimate with high probability. More precisely, given any unbiased estimator X_A of per(A), the number of samples needed to approximate per(A) within a factor $1 \pm \varepsilon$ with high probability is $\frac{\mathrm{const}}{\varepsilon^2} \frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2}$. We call the ratio $\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2}$ of the second moment to the square of the expectation the *critical ratio* of the estimator.

Karmarkar *et al.* [12] showed that the critical ratio of Godsil and Gutman's estimator when run on any $n \times n$ $(0,1)$ matrix A is bounded above by $3^{n/2}$. More remarkably, they also showed that if each 1-entry of A is replaced not by $\{\pm 1\}$ but by a random element of $\{\pm 1, \pm i\}$ (where i is the complex square root of -1),† forming a matrix C, then the analogous estimator $|\det(C)|^2$ is still unbiased and the bound on the critical ratio drops to $2^{n/2}$. This is still exponential, but substantially smaller than the Godsil-Gutman version. In addition, Frieze and Jerrum [7] showed that the critical

†Actually Karmarkar *et al.* used complex *cube* roots of unity, rather than fourth roots as stated here. We prefer the latter choice as it fits more naturally into our generalized framework. It is not hard to check that the use of kth roots for any $k \geq 3$ leads to essentially the same asymptotic behavior.

ratio of the Karmarkar *et al.* estimator is polynomially bounded with high probability for a *random* $(0,1)$ matrix A.

These ideas were pushed further by Barvinok [3] in a rather different framework. Instead of asking how much time is needed to compute a $(1 \pm \varepsilon)$ approximation of $\mathrm{per}(A)$, Barvinok asked how good an approximation can be obtained in polynomial time. Under this measure, he showed that the original Godsil-Gutman idea could also be improved by replacing each 1-entry of A by a continuous sample from a standard normal distribution; the resulting algorithm approximates $\mathrm{per}(A)$ within a factor of about $(3.57)^n$ with high probability in polynomial time. Moreover, the extension of Barvinok's algorithm to the complex numbers, analogous to that of Karmarkar *et al.*, provides an improvement of this approximation ratio to about $(1.79)^n$. Barvinok also showed that a further extension to the *quaternion* algebra (i.e., each 1-entry of A is replaced by a value $b_1 + b_2 i + b_3 j + b_4 k$, where i, j, k are Hamilton's quaternions and the b's are independent standard normal) again improves the approximation ratio to about $(1.32)^n$. Finally, Barvinok observed that $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ are the first three *Clifford algebras* and suggested that it might be possible to generalize his estimators to these. In a subsequent paper [4], Barvinok also proposed an extension of his techniques to higher dimensional real matrix algebras and conjectured that, for sufficiently high dimension, it may yield a polynomial time approximation algorithm for the permanent within ratio c^n for any desired constant $c > 1$. (Note that this is a much weaker requirement than that of a fully polynomial randomized approximation scheme.)

In this paper, we apply higher dimensional Clifford algebras to the original approximation scheme framework, and demonstrate a more dramatic improvement than that conjectured by Barvinok. We begin with the observation that the sets $\{1\}$, $\{1, i\}$, $\{1, i, j, k\}$ are the basis elements of the first three Clifford algebras [‡] of dimensions 1, 2 and 4 respectively. For each $m \geq 1$, we define a permanent estimator based on the Clifford algebra CL_m of dimension 2^{m-1}. The estimator is analogous to that of Godsil-Gutman and Karmarkar *et al.*, and is very easy to describe: simply replace each 1-entry of A by an independent element chosen u.a.r. from $\{\pm e_1, \pm e_2, \ldots, \pm e_{2^{m-1}}\}$, where the e_i are the basis elements of CL_m, to obtain a matrix B; then output $|\det(B)|^2$. Note that $\det(B)$ is a value in CL_m; the norm-square function $|\cdot|^2$ is simply the sum of squares of the coefficients in the above basis. It is not hard to show that this estimator remains unbiased for all m.

Our first main result is that the critical ratio of the estimator decreases dramatically with the dimension. Specifically, we show:

Theorem A *Let* $X_A = |\det(B)|^2$ *be the value output by the above algorithm over* CL_m*, with* $m = 4q+2$*. Then* $\mathrm{E}[X_A] = \mathrm{per}(A)$ *and the critical ratio* $\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2}$ *for any* $n \times n$ *matrix* A *is bounded above by* $(1 + \frac{1}{2^{2q}})^{n/2}$.

An immediate corollary of this theorem is that, if we put $q = \lceil \frac{1}{2} \log_2 n \rceil$, then the critical ratio is bounded by a constant! — i.e., a constant number of trials suffice to get a good approximation of $\mathrm{per}(A)$. Moreover, to achieve this we need only work in the algebra CL_{4q+2} of dimension $2^{4q+1} = O(n^2)$, which is also polynomial in n. Thus, in principle, the approach yields a fully-polynomial randomized approximation scheme for the permanent.[§]

The only catch is that our estimator requires the computation of $|\det(B)|^2$, where B is a matrix of basis elements of a high-dimensional algebra. The algebras CL_m are *not commutative* for $m \geq 3$ (CL_1 is the reals; CL_2 is the complex numbers; CL_3 is the quaternions), so standard polynomial time determinant computations break down. (In fact, there is evidence that computing general determinants in a non-commutative setting is computationally infeasible [17].) Nonetheless, we are able to overcome this obstacle at least for the first interesting case, namely the quaternions; for this

[‡]More accurately, the *second* Clifford algebras. For definitions see the next section.

[§]We have chosen the values $m \equiv 2 \bmod 4$ to allow the cleanest statement of Theorem A. In fact the critical ratio is monotonically decreasing with m, and our techniques allow a similar bound to be computed for any m.

algebra, our general analysis shows that the critical ratio is at most $(3/2)^{n/2}$. In our second main result, we show how to define a modified quaternion estimator closely related to the original one, but easily computable in polynomial time.¶ Surprisingly, we show that this modified estimator has the same first and second moments as the original one, yielding:

Theorem B *There is a quaternion-based unbiased estimator for the permanent that is computable in polynomial time and has critical ratio at most* $(3/2)^{n/2}$.

Recall that the estimator of Karmarkar *et al.* has critical ratio $2^{n/2}$, so Theorem B gives a further significant exponential improvement. We leave as an intriguing open problem the question of whether the higher-dimensional estimators with small variance whose existence is guaranteed by Theorem A also have modified versions that are computable in polynomial time.

The following is a brief road-map of the paper. In section 2 we present the minimal set of facts about Clifford algebras required to understand our work. We go on in section 3 to define our generalized estimators based on the Clifford algebras CL_m, and show that they are always unbiased. The bulk of this section is devoted to the proof of the bound on the critical ratio of these estimators, as stated in Theorem A. The proof exploits the substantial group-theoretic structure underlying the Clifford algebras, and offers as a byproduct new insights into why the introduction of complex numbers by Karmarkar *et al.* improves on the initial Godsil-Gutman algorithm. In section 4 we define and analyze the modified estimator over the quaternion algebra, thus proving Theorem B.

2 Clifford Algebras

In this section we cover the necessary fundamentals of Clifford algebras that we require for our estimators. There is a great deal of theory on Clifford algebras, but we will present only the minimal required background. For further reading we recommend, e.g., [14].

The Clifford algebras we will use are real algebras with basis elements of the form $u_{a_1 a_2 \dots a_k}$, with $a_i \in [m] = \{1, \dots, m\}$ and $a_i < a_{i+1}$, together with $u_\epsilon = 1$. The multiplication rules are simple: for $i \neq j \in [m]$, $u_i u_j = u_{ij} = -u_j u_i$, and $u_i u_i = 1$. A general element of the Clifford algebra can thus be written as $h = \sum_S c_S u_S$, where c_S is real and S ranges over subsets of $[m]$.

We will restrict ourselves to the "second Clifford algebras," a set of subalgebras of the full Clifford algebras defined above. In a second Clifford algebra, every basis element must have an even number of subscripts: e.g., u_{12}, u_{2467}, etc. We denote the mth such Clifford algebra CL_m. Clearly the number of basis elements of CL_m is the number of even-cardinality subsets of $[m]$, which is 2^{m-1}. Of course this is also the dimension of the algebra over the reals.

The first few Clifford algebras are familiar enough. CL_1 has 1 as its only basis element and is just the real numbers, $\mathbb{R}$. CL_2 has basis $\{1, u_{12}\}$, and is in fact the complex numbers, $\mathbb{C}$, with $i = u_{12}$. (Note that $u_{12}^2 = u_1 u_2 u_1 u_2 = -u_1^2 u_2^2 = -1$.) CL_3 has basis $\{1, u_{12}, u_{23}, u_{13}\}$ and is the quaternions, $\mathbb{H}$, with $u_{23} = j$ and $u_{13} = k$. (The reader may check the familiar properties $i^2 = j^2 = k^2 = -1$ and $ij = k = -ji$.) Note that CL_3 (and hence CL_m for all $m \geq 3$) is not commutative; however, if two basis elements do not commute then their two products differ only up to a sign. The reader may consult the multiplication table for CL_4 in Appendix A.

Conjugation in CL_m is defined in the natural way. The conjugate of a basis element u_S, written $\overline{u_S}$, is its inverse, i.e., the (unique) element that satisfies $u_S \overline{u_S} = \overline{u_S} u_S = 1$. The conjugate of a general element $u = \sum_S c_S u_S$ is defined as $\overline{u} = \sum_S c_S \overline{u_S}$. Note that in general we cannot construct the inverse of u from $\overline{u}$, as $u\overline{u}$ may not be real; indeed, in CL_m for $m \geq 4$ not every element has an inverse.

The following useful observations can readily be verified from the above definitions:

¶The determinant computation is a discrete version of Barvinok's quaternion estimator [3], but the analysis is substantially different in that we are analyzing the second moment while he was analyzing the tails.

1. A basis element u_S is *self-conjugate* (i.e., u_S is its own inverse, $u_S = \overline{u_S}$) if and only if S consists of $4k$ subscripts. If S consists of $4k+2$ subscripts, then $\overline{u_S} = -u_S$. Notice that for any S, $u_S^2 = \pm 1$.

2. The product of two basis elements u_S and $u_{S'}$ is $\epsilon u_{S \oplus S'}$, where $S \oplus S'$ is the disjoint union of S and S' and ϵ is a sign that depends on the number of inversions necessary to produce $S \oplus S'$ from S and S'.

3. Two basis elements commute if and only if the number of subscripts they share is even; i.e., $u_S u_{S'} = u_{S'} u_S$ if and only if $|S \cap S'| = 2k$. Otherwise, $u_S u_{S'} = -u_{S'} u_S$.

4. If u, u' are signed basis elements chosen independently and uniformly at random, then their product uu' is also a uniformly random signed basis element.

It will be convenient to associate with each Clifford algebra CL_m the group of its 2^m *signed* basis elements G_m. Each group element $\alpha \in G_m$ corresponds to either u_S or $-u_S$ for some $S \subseteq [m]$, and the group operation is simply multiplication as defined in CL_m. For $\alpha \in G_m$, we denote by u_α the corresponding signed basis element in CL_m. Thus an arbitrary element of CL_m can be written as $h = \sum_{\alpha \in G_m} c_\alpha u_\alpha$ for $c_\alpha \in \mathbb{R}^+$. We assume that c_α is non-zero for at most one of u_S and $-u_S$, so that this representation is unique.

Recall from the introduction that our permanent estimators are of the form $|\det(B)|^2$, where the entries of the matrix B, and therefore also $\det(B)$, lie in the Clifford algebra CL_m. Thus we need to define the norm-square $|u|^2$ for $u \in CL_m$. Generalizing from the reals, complex numbers and quaternions, we might try to use the definition $|u|^2 = u\overline{u}$. However, this is problematic in CL_m for $m \geq 4$ because $u\overline{u}$ is not in general real. Instead, we will define $|u|^2$ to be the *real part* of $u\overline{u}$; equivalently, if $u = \sum_S c_S u_S$, then $|u|^2 = \sum_S c_S^2$ (which is just the squared length of u, viewed as a 2^{m-1}-dimensional vector).

Because the Clifford algebras are non-commutative, we also need to exercise care in our definition of determinant. Many alternative definitions of determinant that are equivalent for matrices over a field turn out not to be equivalent in the absence of commutativity. This topic has been studied before for general non-commutative rings (see, e.g. [8]), and particularly in the context of the quaternions (see [2] for an overview). Unless otherwise stated, we will define determinant as $\det(B) = \sum_{\pi \in S_n} \mathrm{sgn}(\pi) \prod_{i=1}^n b_{i,\pi i} = \sum_{\pi \in S_n} \mathrm{sgn}(\pi) b_{1,\pi 1} \cdots b_{n,\pi n}$. (Note that in a non-commutative setting the order of multiplication in each monomial is significant; we take it to be in row order.) This definition is sometimes called the "Cayley determinant" or "row-determinant." Another closely related definition of interest is the "Moore determinant," which is the same as the Cayley determinant except that the order of the terms in each monomial is determined by the cycle structure of the corresponding permutation. Finally there are two definitions that apply to the quaternions but not immediately to higher-dimensional Clifford algebras: the "Dieudonné determinant," which is based on Gaussian elimination, and the "Study determinant," which relates a quaternion matrix to a larger complex matrix. These latter two notions are closely related and we will use both in the implementation of our quaternion estimator in Section 4.

3 General Clifford Algebra Estimators

3.1 Definition and expectation

For each Clifford algebra CL_m, we can define a corresponding estimator for the permanent in the natural way: given a $(0,1)$ matrix A, replace each 1-entry of A with a signed basis element of CL_m chosen independently and uniformly at random to obtain a new matrix B; then compute $X_A = |\det(B)|^2$, where $\det(B)$ is the Cayley determinant defined above, which lies in CL_m; and $|\det(B)|^2$ is the real part of $\det(B)\,\overline{\det(B)}$. We prove first that this estimator is unbiased for all m.

The proof is similar to the proofs for the low-dimensional versions of Godsil-Gutman ($m = 1$) and Karmarkar *et al.* ($m = 2$), but is complicated by the fact that $u\overline{u}$ is not necessarily real.

Proposition 3.1 *In any Clifford algebra CL_m, we have* $\mathrm{E}[X_A] = \mathrm{per}(A)$.

Proof: We first introduce some notation. Given a permutation π, we define B_π to be $\prod_{i=1}^{n} b_{i,\pi i}$ and $\overline{B}_\pi$ to be $\prod_{i=n}^{1} \overline{b}_{i,\pi i}$. Thus $B_\pi \overline{B}_\pi = \prod_{i=1}^{n} a_{i,\pi i}$. Further, given two permutations π_1 and π_2, we say that $\mathrm{R}(B, \pi_1, \pi_2) = 1$ if $B_{\pi_1}\overline{B}_{\pi_2}$ is real and 0 otherwise. Note that $\mathrm{R}(B, \pi, \pi) = 1$ for all π.

We can write $\mathrm{E}[X_A] = \sum_B \Pr(B) \sum_{\pi_1\pi_2} \mathrm{sgn}(\pi_1\pi_2) B_{\pi_1}\overline{B}_{\pi_2}\mathrm{R}(B, \pi_1, \pi_2)$ where the sum is over all possible choices of the random matrix B and $\Pr(B)$ is the probability of choosing B. We then proceed as follows:

$$
\begin{aligned}
\mathrm{E}[X_A] &= \sum_{\pi_1}\sum_{B} \Pr(B) B_{\pi_1}\overline{B}_{\pi_1}\mathrm{R}(B, \pi_1, \pi_1) + \sum_{\pi_1\neq\pi_2}\sum_{B} \Pr(B) B_{\pi_1}\overline{B}_{\pi_2}\mathrm{R}(B, \pi_1, \pi_2) \\
&= \sum_{\pi_1}\prod_{i} a_{i,\pi i} + \sum_{\pi_1\neq\pi_2}\sum_{B} \Pr(B) B_{\pi_1}\overline{B}_{\pi_2}\mathrm{R}(B, \pi_1, \pi_2) \\
&= \mathrm{per}(A) + \sum_{\pi_1\neq\pi_2}\sum_{B} \Pr(B) B_{\pi_1}\overline{B}_{\pi_2}\mathrm{R}(B, \pi_1, \pi_2).
\end{aligned}
$$

To finish the proof, note that all choices of B are equally likely. When $\pi_1 \neq \pi_2$, let j be the smallest index such that $\pi_1 j \neq \pi_2 j$. Then $b_{j,\pi_1 j}$ is chosen independently of the other basis elements, and for each value u_S it takes on, it takes on $-u_S$ with equal probability; and in each case the value of $R(B, \pi_1, \pi_2)$ is the same. Thus the sum on the right is 0 and $\mathrm{E}[X_A] = \mathrm{per}(A)$. $\square$

3.2 The second moment: block diagonal case

Recall that the efficiency of the estimator X_A is governed by its critical ratio, $\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2}$. Thus we need to compute the second moment, $\mathrm{E}[X_A^2]$. We first perform a detailed analysis for block diagonal matrices, and then in the next subsection use this to derive a bound for all matrices. Let $A_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Then the block diagonal matrix A_n is the $2n \times 2n$ matrix with n copies of A_1 along its diagonal. From Proposition 3.1 we have $\mathrm{E}[X_{A_n}] = \mathrm{per}(A_n) = 2^n$. For convenience we define A_0 to be the 1×1 identity matrix; note that X_{A_0} is identically 1. We will study the distribution of X_{A_n} in the algebra CL_m as m varies.

The main result of this section is the following theorem:

Theorem 3.2 *Let A_n be the block diagonal matrix defined above, and let $m = 4q+2$ for some $q \in \mathbb{N}$. Then in CL_m, we have that* $\mathrm{E}[X_{A_n}^2] \leq [4(1+\frac{1}{2^{m/2}})]^n$, *and thus the critical ratio* $\frac{\mathrm{E}[X_{A_n}^2]}{\mathrm{E}[X_{A_n}]^2} \leq (1+\frac{1}{2^{m/2}})^n$.

Before embarking on the proof of the theorem, we first provide some geometric intuition as to why increasing the dimension of the Clifford algebras decreases the variance of our estimator. Let $B_1 = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ denote the random matrix computed by the algorithm when run on A_1. Then $\det(B_1) = b_{11}b_{22} - b_{12}b_{21}$. Since the product of two random signed basis elements in CL_m is again a random signed basis element, it is clear that $\det(B_1)$ has the same distribution as $a + b$, where a, b are independent random signed basis elements. Thus, in distribution, $X_{A_1} = |a + b|^2$.

Now, if we consider CL_m as a 2^{m-1}-dimensional vector space over $\mathbb{R}$, then the basis elements a and b are signed elementary basis vectors, and $X_{A_1} = |a + b|^2$ is the squared length of the vector $a + b$. The estimator is exact if a and b are orthogonal, in which case $X_{A_1} = 2 = \mathrm{per}(A_1)$. The variance results entirely from the remaining cases $a = \pm b$, which give $X_{A_1} = 4$ and $X_{A_1} = 0$, respectively. Since a and b are randomly chosen from a set of size 2^m, we have in particular that

$$X_{A_1} = \begin{cases} 4 & \text{if } a = b \quad \text{(with probability } 1/2^m\text{);} \\ 0 & \text{if } a = -b \quad \text{(with probability } 1/2^m\text{);} \\ 2 & \text{otherwise} \quad \text{(with probability } 1 - 2/2^m\text{)} \end{cases}$$

and therefore that the second moment is

$$\mathrm{E}[X_{A_1}^2] = \left(16 \times \tfrac{1}{2^m}\right) + \left(0 \times \tfrac{1}{2^m}\right) + \left(4 \times (1 - \tfrac{2}{2^m})\right) = 4\left(1 + \tfrac{1}{2^{m-1}}\right).$$

Observe how increasing the dimension of CL_m increases the probability that two randomly chosen elementary signed basis vectors are orthogonal, which in turn decreases the variance of X_{A_1}.

We now proceed with the proof of Theorem 3.2 for general n. Let B_n denote the random matrix computed by the estimator when run on the block-diagonal matrix A_n. Then, in distribution, $\det(B_n) = (a_1 + b_1)(a_2 + b_2) \cdots (a_n + b_n)$ for mutually independent random basis elements a_i and b_i. We wish to compute the second moment of $X_{A_n} = |\det(B_n)|^2$.

For the first three values of m, namely $CL_1 = \mathbb{R}$, $CL_2 = \mathbb{C}$, and $CL_3 = \mathbb{H}$, it is trivial to extend the above analysis for $n = 1$ to general n. In these three algebras the norm-squared function $|\cdot|^2$ is real-valued and a multiplicative homomorphism, so we have $X_{A_n} = \prod_i |a_i + b_i|^2$, and $\mathrm{E}[X_{A_n}^2] = \mathrm{E}[X_{A_1}^2]^n = [4(1 + \frac{1}{2^{m-1}})]^n$. However, for the general case $m \geq 4$ we need to do some work since $|\cdot|^2$ is no longer a homomorphism. Our argument will exploit the detailed structure of the algebras CL_m.

The proof of Theorem 3.2 for general n and q is contained in Lemmas 3.3, 3.4, and 3.5 below. The proofs of these lemmas in turn depend on the following basic observation. Recall that, in distribution, $\det(B_n) = (a_1 + b_1)(a_2 + b_2) \cdots (a_n + b_n)$ for mutually independent random signed basis elements a_i and b_i. Therefore, still in distribution, $\det(B_n)\overline{\det(B_n)} = (a_1 + b_1) \cdots (a_n + b_n)(\overline{a_n} + \overline{b_n}) \cdots (\overline{a_1} + \overline{b_1}) = (a_1 + b_1)\det(B_{n-1})\overline{\det(B_{n-1})}(\overline{a_1} + \overline{b_1})$. Rewriting slightly, we get

$$\det(B_{n+1})\overline{\det(B_{n+1})} = a(1 + c)\det(B_n)\overline{\det(B_n)}(1 + \overline{c})\overline{a}, \tag{1}$$

where, again, a and c are independent random signed basis elements.

The success of the Clifford algebra estimators can be explained by the algebraic restrictions on the behavior of X_{A_n}. The concrete ideas are contained in the following lemma and its proof:

Lemma 3.3 *Suppose we are working in CL_m. Then, for any choice of B_n, $\det(B_n)\overline{\det(B_n)}$ is either zero or of the form $2^k \sum_{\alpha \in G} u_\alpha$, where k is a non-negative integer and G is a self-conjugate subgroup of G_m (i.e., $u_\alpha^2 = 1$ for all $\alpha \in G$).*

Note that any self-conjugate subgroup G is necessarily abelian. Note also that the above representation assumes that G does not contain both α and $-\alpha$; all the self-conjugate subgroups in the sequel can easily be seen to have this property and we will assume it from now on. We will write $-G$ to denote the set of $\alpha \in G_m$ such that $-\alpha \in G$.

Proof of Lemma 3.3 (sketch): We proceed by induction on n. In the base case $n = 0$, we always have $\det(B_0)\overline{\det(B_0)} = 1$, which can be written in the form $2^k \sum_{\alpha \in G} u_\alpha$ with $k = 0$ and G the trivial subgroup $\{1\}$. Now, applying the induction hypothesis to (1) and expanding, we get

$$\det(B_{n+1})\overline{\det(B_{n+1})} = 2^k a\Big(\sum_{\alpha \in G} u_\alpha + c(\sum_{\alpha \in G} u_\alpha)\overline{c} + c(\sum_{\alpha \in G} u_\alpha) + (\sum_{\alpha \in G} u_\alpha)\overline{c}\Big)\overline{a}. \tag{2}$$

Our task is therefore to show that, for an arbitrary self-conjugate subgroup G and signed basis elements a and c, the r.h.s of (2) is either zero or can be written in the form $2^{k'} \sum_{\alpha \in G'} u_\alpha$ for some

$k' \geq 0$ and self-conjugate subgroup G'. This follows from a fairly straightforward case analysis, whose structure we outline below. The proofs for each case are deferred to Appendix B. (Cases 1 and 2 are analogous to the cases $a = \pm b$ and a orthogonal to b, respectively, in our earlier analysis for the case $n = 1$.)

Case 1: $c \in G \cup -G$.

Here if $c \in G$ then the outside coefficient 2^k quadruples to 2^{k+2}, and the subgroup G *transmutes* to another subgroup $G' = aG\overline{a}$ of the same size as G. Otherwise, if $c \in -G$, we obtain 0.

Case 2: $c \notin G \cup -G$.

We divide this case into two more subcases.

case 2a: c commutes with G, $c = \overline{c}$. Here the coefficient 2^k doubles to 2^{k+1}, while G *expands* to a subgroup $G' = a(G \cup cG)\overline{a}$ of twice the size of G.

case 2b: All other cases (c commutes with G and $c = -\overline{c}$; or c does not commute with G). Here the coefficient doubles to 2^{k+1} while G transmutes to another subgroup G' of the same size.

This completes the sketch of the proof of Lemma 3.3. □

Example: We illustrate each of the above cases with a running example from CL_5, the 16-dimensional algebra whose subscripts are drawn from $\{1, \ldots, 5\}$. We start with $\det(B_0)\overline{\det(B_0)} = 1$ (so $k = 0$ and G is the trivial subgroup $\{1\}$) and follow the evolution of $\det(B_n)\overline{\det(B_n)}$ for some particular sequence of choices of c. (We will fix $a = 1$ throughout for ease of computation.)

- $c = u_{1234}$ (case 2a). Expression (1) is $(1+u_{1234})(1)(1+u_{1234})$, which simplifies to $2(1+u_{1234})$. The outside coefficient has doubled to 2 and G has expanded to the subgroup $\{1, u_{1234}\}$.

- $c = u_{1234}$ (case 1). Here expresssion (1) is $2(1 + u_{1234})(1 + u_{1234})(1 + u_{1234})$, which simplifies to $8(1 + u_{1234})$. The outside coefficient has quadrupled to 8 and G has transmuted (to itself, because $a = 1$). (Note that if c had been $-u_{1234}$, we would have obtained 0.)

- $c = u_{23}$ (case 2b). Here expression (1) is $8(1 + u_{23})(1 + u_{1234})(1 - u_{23})$, which simplifies to $16(1 + u_{1234})$. The outside coefficient has doubled and G has transmuted (to itself).

- $c = u_{25}$ (case 2b). Here expression (1) is $16(1 + u_{25})(1 + u_{1234})(1 - u_{25})$, which simplifies to $32(1 - u_{1345})$. The outside coefficient has doubled and G has transmuted to the subgroup $\{1, -u_{1345}\}$. □

The proof of Lemma 3.3 reveals a simple pattern to the behavior of $\det(B_n)\overline{\det(B_n)}$ that allows us to easily bound $\mathrm{E}[X^2_{A_n}]$. Note that since X_{A_n} is the real part of $\det(B_n)\overline{\det(B_n)}$, its value is just the outside coefficient 2^k.

Lemma 3.4 *Let p be the maximum possible value of the ratio $2|G|/|G_m|$ over all self-conjugate subgroups G in G_m. Then in CL_m we have $\mathrm{E}[X^2_{A_n}] \leq [4(1+p)]^n$, and thus the critical ratio satisfies $\frac{\mathrm{E}[X^2_{A_n}]}{\mathrm{E}[X_{A_n}]^2} \leq (1 + p)^n$.*

Proof: We again use induction on n. In the base case $n = 0$, we have $X_{A_0} = 1$ with probability 1, and hence $\mathrm{E}[X^2_{A_0}] = 1$. For the inductive step we examine the random variable $\det(B_{n+1})\overline{\det(B_{n+1})}$. Recall that, conditioned on the value of $\det(B_n)\overline{\det(B_n)}$, the distribution of this r.v. is as in (1) where a, c are independent random signed basis elements. From Lemma 3.3 we know that $\det(B_n)\overline{\det(B_n)}$ is either zero or of the form $2^k \sum_{\alpha \in G} u_\alpha$ for some k and G; thus the r.v. X_{A_n} has value zero or 2^k respectively. From the proof of Lemma 3.3 we see that the outside coefficient 2^k exactly doubles in all cases except case 1. In this latter case it either quadruples (if $c \in G$) or becomes zero (if $c \in -G$). Plainly each of these outcomes occurs with probability $|G|/|G_m|$. Thus, conditioned on X_{A_n}, the distribution of $X_{A_{n+1}}$ is

$$\begin{cases} 4X_{A_n} & \text{with probability } |G|/|G_m|; \\ 0 & \text{with probability } |G|/|G_m|; \\ 2X_{A_n} & \text{with probability } 1 - 2|G|/|G_m|. \end{cases}$$

Since $|G|/|G_m| \le p/2$ by definition of p, we therefore have

$$\mathrm{E}[X^2_{A_{n+1}}] \le (16\tfrac{p}{2} + 4(1-p))\mathrm{E}[X^2_{A_n}] = 4(1+p)\mathrm{E}[X^2_{A_n}]. \tag{3}$$

This completes the proof by induction on n. □

Lemma 3.4 bounds the second moment in terms of $p = \max \frac{2|G|}{|G_m|}$, where the maximum is over all self-conjugate subgroups G in CL_m. The final ingredient is to show that p decreases rapidly as a function of m:

Lemma 3.5 *Let $m = 4q+2$, and let p be defined as above. Then in CL_m, $p \le \frac{1}{2^{2q+1}}$.*

Proof of Lemma 3.5 (sketch): We shall give a very simple argument that yields a slightly weaker bound, namely $p \le \frac{1}{2^q}$, and conveys the main idea. The additional factor of 2 in the exponent requires some slightly more detailed analysis (see Proposition B.9 in Appendix B).

Let G be a self-conjugate subgroup of G_m, and let H be the subgroup $G \cap G_{m-1}$. It is easy to check that either $H = G$ or $|H| = |G|/2$. This tells us that for any self-conjugate subgroup G of G_m, there is a subgroup of at least half its size in G_{m-1}. Hence p does not increase as m increases.

Now consider CL_{4q}. The basis element g that contains every index (e.g., $u_{12345678}$ in CL_8) is self-conjugate and commutes with every other basis element. This element (or its negation) must be contained in every maximal self-conjugate subgroup G of G_{4q}; otherwise $G \cup gG$ is a larger such subgroup.

Moving to CL_{4q+1}, we wish to show that the size of any maximal self-conjugate subgroup $G \subseteq G_{4q+1}$ is unchanged from the size of a maximal self-conjugate subgroup in G_{4q}, and hence p decreases by a factor of 2. Consider $G' = G \cap G_{4q}$, a self-conjugate subgroup of G_{4q}. If $G' = G$, then G is already a subgroup of G_{4q} and thus no larger than a maximal self-conjugate subgroup of G_{4q}. Otherwise G' is smaller than G and therefore exactly half the size of G. We now observe that G' cannot be a maximal self-conjugate subgroup of G_{4q}: if it were, it would contain one of the elements $\pm g$, but no element of $G_{4q+1} - G_{4q}$ commutes with g. Thus G' is half the size of a maximal self-conjugate subgroup, and G is the same size as a maximal subgroup. □

Putting Lemmas 3.4 and 3.5 together, we are done with the proof of Theorem 3.2 and hence with the main business of this section. Theorem 3.2 and its proof contain the essential intuition about the behavior of the second moment on block diagonal matrices as m increases, and immediately imply that for $m = O(\log n)$, the critical ratio is bounded above by a constant. For technical reasons, in order to bootstrap this to a bound for general matrices we actually need to derive the exact form of $\mathrm{E}[X^2_{A_n}]$ as a sum of exponentials $\sum_i c_i E_i^n$, not just an upper bound $[4(1+\frac{1}{2^{2q+1}})]^n$ as in Theorem 3.2. Somewhat remarkably, it turns out that $\mathrm{E}[X^2_{A_n}]$ is the sum of only two exponentials, as the following theorem states. The proof follows from a slightly more refined analysis of the behavior of the subgroups than that used in the proof of Lemma 3.4, and may be found in Appendix B.

Theorem 3.6 *Let $m = 4q+2$. In CL_m, $\mathrm{E}[X^2_{A_n}] = c_1 E_1^n + c_2 E_2^n$, where $E_1 = 4(1+\frac{1}{2^{2q+1}})$, $E_2 = 4(1-\frac{1}{2^{2q+1}})$, and c_1, c_2 are non-negative constants with $c_1 + c_2 = 1$. Thus in particular $\mathrm{E}[X^2_{A_n}] \le E_1^n$.*

Remark: The reason we choose $m \equiv 2 \bmod 4$ is to allow the cleanest possible formulation of Lemma 3.5 and Theorem 3.6. However, the essential point is that a constant factor increase in the dimension (i.e., a constant additive increase in m) leads to a constant factor decrease in p (the relative size of the maximum subgroup). In fact, we have seen that p is monotonically decreasing with m, and a more detailed analysis shows that p decreases by a factor of 2 when $m \equiv 1, 2, 3$ or $5 \bmod 8$, and otherwise remains unchanged. □

3.3 The second moment: general case

We now extend our bound from the block diagonal case to general matrices. This analysis will rely on the exponential form of the second moment in the block diagonal case from Theorem 3.6.

Theorem 3.7 *Suppose that in CL_m, the second moment for the block diagonal matrix A_n satisfies $\mathrm{E}[X_{A_n}^2] = c_1 E_1^n + c_2 E_2^n$ for $2 \le E_2 < E_1$ and non-negative constants c_1, c_2 with $c_1 + c_2 = 1$. Then the critical ratio for an arbitrary $n \times n$ matrix A satisfies*

$$\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2} \le \sum_{i=1}^{2} c_i \left(\frac{E_i - 2}{2}\right)^{n/2} \le \left(\frac{E_1 - 2}{2}\right)^{n/2}.$$

Substituting the value $E_1 = 4(1 + \frac{1}{2^{2q+1}})$ from Theorem 3.6 immediately yields Theorem A in the Introduction.

Before embarking on the proof of Theorem 3.7, we introduce a useful graph-theoretic framework from [12]. Recall that we may view $\mathrm{per}(A)$ as the number of perfect matchings in a bipartite graph $\mathcal{B}(A)$ in which (i,j) is an edge if and only if $a_{ij} = 1$. We define $P(A)$ to be the set of permutations $\{\pi \in S_n : a_{i,\pi i} = 1\, \forall i\}$, which correspond naturally to perfect matchings in $\mathcal{B}(A)$ (and we will blur this distinction). Thus $\mathrm{per}(A) = |P(A)|$. We need the following observations:

(i) Given two perfect matchings π_1 and π_2, let the graph G be their union $\pi_1 \cup \pi_2$. Then G is a disjoint union of even-length cycles and isolated edges. We will define $c(G)$ to be the number of cycles of G. Conversely, the cycle cover G can be described as the union of a pair of perfect matchings in $2^{c(G)}$ distinct ways. We write $\mathcal{G}(A)$ as the set of all such cycle covers. We therefore can write $\mathrm{per}(A)^2 = \sum_{G \in \mathcal{G}(A)} 2^{c(G)}$

(ii) Given $G, G' \in \mathcal{G}(A)$, we will say that $G' \subseteq G$ if all of the edges of G' are contained in G, or equivalently, if G' can be formed from G by collapsing some of the cycles of G. Thus there are $\binom{c(G)}{k} 2^k$ graphs $G' \subseteq G$ such that $c(G') = c(G) - k$.

(iii) Consider the union of four perfect matchings $\pi_1, \pi_2, \pi_3, \pi_4$. We will say that $\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4$ is *even* if every edge in the union is covered an even number of times. In this case, $\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4$ forms a cycle cover.

(iv) Consider any $G \in \mathcal{G}(A)$, and let $A(G)$ denote the adjacency matrix of G. Then the estimator $X_{A(G)}$ run on $A(G)$ has the same distribution as $X_{A_{c(G)}}$, the estimator on the block diagonal matrix with $c(G)$ blocks.

Proof of Theorem 3.7: Proceeding as in the proof of Proposition 3.1, we can write

$$\mathrm{E}[X_A^2] = \sum_B \Pr(B) \sum_{\pi_1\pi_2\pi_3\pi_4} \mathrm{sgn}(\pi_1\pi_2\pi_3\pi_4) B_{\pi_1} \overline{B}_{\pi_2} B_{\pi_3} \overline{B}_{\pi_4} \mathrm{R}(B, \pi_1, \pi_2) \mathrm{R}(B, \pi_3, \pi_4) \tag{4}$$

To simplify notation, define $B_{\pi_1\pi_2\pi_3\pi_4} = B_{\pi_1}\overline{B}_{\pi_2}B_{\pi_3}\overline{B}_{\pi_4}$ and write $\mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}]$ to denote the summation $\sum_B \Pr(B) B_{\pi_1\pi_2\pi_3\pi_4} \mathrm{R}(B, \pi_1, \pi_2) \mathrm{R}(B, \pi_3, \pi_4)$.

Our first observation is that $\mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}] = 0$ unless $\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4$ is even. This follows because of the presence, in non-even cases, of an independent factor b in $B_{\pi_1\pi_2\pi_3\pi_4}$ that takes on values $\pm u_S$ with equal probability. Thus we may rewrite equation (4) as

$$\mathrm{E}[X_A^2] = \sum_{G \in \mathcal{G}(A)} \sum_{\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 = G} \mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}], \tag{5}$$

where we can ignore $\mathrm{sgn}(\pi_1\pi_2\pi_3\pi_4)$ since it must be 1 when $\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4$ is even.

We now prove, for any fixed $G \in \mathcal{G}(A)$, that

$$\sum_{\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 = G} \mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}] = \sum_{i=1}^{2} c_i (E_i - 2)^{c(G)}. \tag{6}$$

This is done by induction on $c(G)$. The base case $c(G) = 0$ is verified by noticing that $\pi_1 = \pi_2 = \pi_3 = \pi_4$ must be the same permutation, so $B_{\pi_1\pi_2\pi_3\pi_4} = 1$, and both evaluations of $\mathrm{R}(\cdot)$ are also 1, so the left-hand side of (6) is $1 = \sum_i c_i (E_i - 2)^0$.

Now for any fixed G, let us define $A(G)$ as the $(0,1)$ matrix associated with G. Then

$$\mathrm{E}[X_{A(G)}^2] = \sum_{G' \in \mathcal{G}(A(G))} \sum_{\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 = G'} \mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}].$$

Since one possible instance of G' is G itself, we can rewrite this as

$$\begin{aligned}
\sum_{\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 = G} \mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}] &= \mathrm{E}[X_{A(G)}^2] - \sum_{G' \subset G} \sum_{\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 = G'} \mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}] \\
&= \sum_i c_i E_i^{c(G)} - \sum_{G' \subset G} \sum_{\pi_1 \cup \pi_2 \cup \pi_3 \cup \pi_4 = G'} \mathrm{RE}[B_{\pi_1\pi_2\pi_3\pi_4}] \\
&= \sum_i c_i E_i^{c(G)} - \sum_{k=1}^{c(G)} \tbinom{c(G)}{k} 2^k \sum_i c_i (E_i - 2)^{c(G)-k} \\
&= \sum_i c_i E_i^{c(G)} - \sum_i c_i (E_i^{c(G)} - (E_i - 2)^{c(G)}) \\
&= \sum_i c_i (E_i - 2)^{c(G)}
\end{aligned}$$

In the second line here we have used Theorem 3.6 together with observation (iv) from earlier; in the third line we have used the induction hypothesis and observation (ii).

This completes the inductive proof of (6). Plugging the result into equation (5) gives

$$\mathrm{E}[X_A^2] = \sum_{G \in \mathcal{G}(A)} \sum_{i=1}^{2} c_i (E_i - 2)^{c(G)} \le \sum_{G \in \mathcal{G}(A)} (E_1 - 2)^{c(G)}.$$

Finally, combining this with the observation that $\mathrm{E}[X_A]^2 = \sum_{G \in \mathcal{G}(A)} 2^{c(G)}$ we obtain

$$\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2} \le \frac{\sum_{G \in \mathcal{G}(A)} (E_1 - 2)^{c(G)}}{\sum_{G \in \mathcal{G}(A)} 2^{c(G)}} \le \max_{G \in \mathcal{G}(A)} \frac{(E_1 - 2)^{c(G)}}{2^{c(G)}} \le \left(\frac{E_1 - 2}{2}\right)^{n/2}.$$

The last inequality above requires the observation that $E_1 \ge 4$ (which follows because $\mathrm{E}[X_{A_n}^2] \ge \mathrm{E}[X_{A_n}]^2 = 4^n$). This completes the proof of the theorem. □

We should note that our analysis includes the real- and complex-based estimators of Godsil-Gutman [6] and Karmarkar *et al.* [12], as well as the quaternion-based estimator as special cases. The second moments of these estimators on the single block matrix A_1 are $\mathrm{E}[X_{A_1}^2] = 8$ for $\mathbb{R}$, $\mathrm{E}[X_{A_1}^2] = 6$ for $\mathbb{C}$, and $\mathrm{E}[X_{A_1}^2] = 5$ for $\mathbb{H}$. Further, since the norm-squared function $|\cdot|^2$ is a homomorphism for these three cases, we can immediately deduce the exact form of the second moments for the block diagonal case as $\mathrm{E}[X_{A_n}^2] = 8^n$ for $\mathbb{R}$, $\mathrm{E}[X_{A_n}^2] = 6^n$ for $\mathbb{C}$, and $\mathrm{E}[X_{A_n}^2] = 5^n$ for $\mathbb{H}$. Applying Theorem 3.7 then gives the following result:

Corollary 3.8 *The critical ratio for the estimator X_A is bounded above by $\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2} \le 3^{n/2}$ for $\mathbb{R}$, $\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2} \le 2^{n/2}$ for $\mathbb{C}$, and $\frac{\mathrm{E}[X_A^2]}{\mathrm{E}[X_A]^2} \le (\frac{3}{2})^{n/2}$ for $\mathbb{H}$.*

For the real and complex cases, these are the same as the bounds derived by less general methods in [12].

4 Computing the estimator

We turn now to the question of implementing the estimators of the previous section. These estimators are defined in terms of the symbolic determinant of a matrix whose entries are basis elements of a high-dimensional Clifford algebra. Since such algebras are non-commutative, it is not clear how to perform such a computation in polynomial time; indeed, it is known that computing general determinants in a non-commutative setting is computationally infeasible [17].

Our goal in this final section is to show that this difficulty *can* be overcome at least in the first interesting case, namely the quaternion algebra $\mathbb{H} = CL_3$. (Recall that this algebra is non-commutative.) What we shall do is to construct a permanent estimator having the same flavor as that of the previous section, but which *is* efficiently computable; although the new estimator will not be equal to the previous one, it will also be unbiased and satisfy the same bound on the critical ratio for the quaternions given in Corollary 3.8. Thus the performance guarantee of the previous section can actually be achieved in polynomial time.

4.1 A modified estimator over the quaternions

Our new estimator Y_A begins as before by replacing each 1-entry of A by a random element from $\mathbb{H}_8 = \{\pm 1, \pm i, \pm j, \pm k\}$, the signed basis elements of the quaternion algebra. Call the resulting random matrix H. Now, however, rather than working with the symbolic determinant $\det(H)$, we use its so-called *Dieudonné determinant*, defined as the result of performing a standard Gaussian elimination procedure on H as follows:

if $n = 1$ then return $\text{Gauss}(H) = h_{11}$
else if column $h_{\cdot 1} = 0$ then return $\text{Gauss}(H) = 0$
else if $h_{11} = 0$ then add any row $h_{i\cdot}$ with $h_{i1} \neq 0$ to row $h_{1\cdot}$
    for all $i > 1$ add row multiple $-h_{i1}h_{11}^{-1}h_{1\cdot}$ to row $h_{i\cdot}$
    return $\text{Gauss}(H) = h_{11}\text{Gauss}(H_{11})$

Note that $\text{Gauss}(H)$ is quaternion-valued. Its value may depend on the row chosen in the third line. However, the classical theory of Dieudonné determinants (see, e.g., [1]) ensures that the norm-square, $|\text{Gauss}(H)|^2$, is well defined (and indeed preserved under any sequence of row and column operations). Note that in the quaternions the norm-square is just $|h|^2 = h\overline{h}$ as this is always real-valued, and hence inverses exist.

Evidently the estimator Y_A can be computed in $O(n^3)$ time. However, despite its resemblance to that of the previous section, it is not at all clear that it inherits the nice properties of that algorithm. Indeed, it is not even clear that it is unbiased. Note in particular that $|\text{Gauss}(H)|^2$ and $|\det(H)|^2$ may differ considerably. For example, if we take $H = \begin{pmatrix} i & j \\ j & i \end{pmatrix}$ then it can easily be checked that $|\text{Gauss}(H)|^2 = 4$ whereas $|\det(H)|^2 = 0$. However, we shall see presently that, when the non-zero entries of H are random quaternion basis elements, then these two quantities share the same first and second moments!

Before proceeding, it will be convenient to note that $|\text{Gauss}(H)|^2$ can be written equivalently as a single complex determinant, known as the "reduced norm", or "Study determinant" of H. This is derived from the representation of the quaternions as 2×2 complex matrices as follows. For $h \in \mathbb{H}$, write h uniquely as $b + cj$, where $b, c \in \mathbb{C}$. Then h is represented by the matrix $\phi(h) = \begin{pmatrix} b & c \\ -\overline{c} & \overline{b} \end{pmatrix}$. Thus in particular the basis elements are represented as

$$\phi(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \phi(i) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}; \quad \phi(j) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad \phi(k) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}. \tag{7}$$

Given an $n \times n$ quaternion matrix H, define a $2n \times 2n$ complex matrix $D = D_H$ by

$$d_{ij} = \left[\phi(h_{\lceil \frac{i}{2} \rceil, \lceil \frac{j}{2} \rceil})\right]_{(i \bmod 2),(j \bmod 2)}.$$

In words, D_H is formed by replacing each entry h_{ij} of H by its corresponding 2×2 complex matrix, and then erasing the boundaries between these matrices.

We define the *reduced norm* of H as $\det(D_H)$. (Since D_H is a complex matrix, this is well-defined and efficiently computable.) The following fact is easy to check (see, e.g., [2]):

Proposition 4.1 *For any $n \times n$ quaternion matrix H, $|\mathrm{Gauss}(H)|^2 = \det(D_H)$.*

Thus in what follows we may think of Y_A as being defined either as $|\mathrm{Gauss}(H)|^2$ or as $\det(D_H)$. This flexibility will prove useful in our analysis.

Our goal is to show that the above permanent estimator Y_A is unbiased and has the same second moment as the quaternion version of the general estimator derived in the previous section. Thus we will prove the following, which is exactly Theorem B of the Introduction.

Theorem 4.2 *For any $n \times n$ $(0,1)$ matrix A, the Dieudonné determinant estimator Y_A satisfies*

$$\mathrm{E}[Y_A] = \mathrm{per}(A) \qquad \text{and} \qquad \frac{\mathrm{E}[Y_A^2]}{\mathrm{E}[Y_A]^2} \leq \left(\frac{3}{2}\right)^{n/2}.$$

Our analysis will proceed along similar lines to that of the previous section, but the Dieudonné determinant will prove a little harder to work with than the symbolic determinant. In section 4.2 we will deal with the expectation, and in section 4.3 with the second moment.

Remark: This result reveals a connection between the Cayley determinant and the Dieudonné and Study determinants of a quaternion matrix. Other such relationships between the Moore determinant and the Dieudonné and Study determinants are described by Aslaksen [2]. □

4.2 Analysis of expectation

For the expectation, it will be useful to work with the reduced norm formulation of Y_A. Let $D = D_H \in \mathbb{C}^{2n \times 2n}$ be the reduced norm matrix computed by the algorithm. Then we have

$$\mathrm{E}[Y_A] = \mathrm{E}[\det(D)] = \mathrm{E}\Big[\sum_{\pi \in S_{2n}} \mathrm{sgn}(\pi) \prod_{i=1}^{2n} d_{i,\pi i}\Big] \stackrel{\mathrm{def}}{=} \sum_{\pi \in S_{2n}} \mathrm{E}[D_\pi].$$

Clearly each entry d_{ij} of D depends on exactly one entry of H, namely $h_{\lceil \frac{i}{2} \rceil, \lceil \frac{j}{2} \rceil}$. Conversely, each entry h_{ij} of H determines four entries of D, namely $d_{2i-1,2j-1}$, $d_{2i-1,2j}$, $d_{2i,2j-1}$ and $d_{2i,2j}$. Moreover, by (7) we can view these four entries as b_{ij}, c_{ij}, $-\bar{c}_{ij}$ and $\bar{b}_{ij}$ respectively, where b_{ij}, c_{ij} are chosen randomly as follows: flip a fair coin. If Heads, choose b_{ij} u.a.r. from $\mathbb{C}_4 = \{\pm 1, \pm i\}$ and set $c_{ij} = 0$; if Tails, set $b_{ij} = 0$ and choose c_{ij} u.a.r. from $\mathbb{C}_4$.

Now let $\pi \in S_{2n}$. The factors $d_{i,\pi i}$ of D_π depend on a set $\pi|_n$ of between n and $2n$ entries of H, viz.

$$\pi|_n = \left\{(\lceil \tfrac{i}{2} \rceil, \lceil \tfrac{\pi i}{2} \rceil) : i \in [2n]\right\} = \left\{(k, \lceil \tfrac{\pi(2k-1)}{2} \rceil), (k, \lceil \tfrac{\pi(2k)}{2} \rceil) : k \in [n]\right\}.$$

The result is an immediate consequence of the following two claims (recall that $P(A)$ is the set of nonzero permutations in A):

Claim 1: Let $\pi \in S_{2n}$. Then $\mathrm{E}[D_\pi] \neq 0 \Rightarrow \pi|_n \in P(A)$.

Claim 2: Let $\sigma \in P(A)$. Then $\mathrm{E}[D_\sigma] \stackrel{\mathrm{def}}{=} \sum_{\pi : \pi|_n = \sigma} \mathrm{E}[D_\pi] = 1$.

The intuition behind these claims is that the only permutations with nonzero expectation in D are those that correspond exactly to nonzero permutations in A, and each such permutation contributes an expected value of 1, thus yielding the permanent.

To prove Claim 1, consider $\pi \in S_{2n}$ with $\mathrm{E}[D_\pi] \neq 0$. Fix any odd $i = 2k-1 \in [2n]$, and assume $\pi i = 2l-1$ is odd (the case of πi even is handled similarly). Thus $d_{i,\pi i} = b_{kl} \in \mathbb{C}_4$, where b_{kl} is the result of the random experiment described earlier. Since b_{kl} has a random sign, this factor cannot be independent of all other factors in D_π. But since the elements of H are all independent, the only other elements of D which are not independent of b_{kl} are $d_{i,\pi i+1} = c_{kl}$, $d_{i+1,\pi i} = -\overline{c}_{kl}$ and $d_{i+1,\pi i+1} = \overline{b}_{kl}$. And since π is a permutation, the only one of these that can be a factor of D_π is $d_{i+1,\pi i+1}$. Hence we must have $\pi(i+1) = \pi i + 1$. This in turn implies that $\lceil \frac{\pi i}{2} \rceil = \lceil \frac{\pi(i+1)}{2} \rceil$, so $\pi|_n$ contains only one entry in the kth row of H, namely $(k, \lceil \frac{\pi(2k-1)}{2} \rceil)$. Since i was arbitrary, we conclude that $\pi|_n \in S_n$. Clearly $\pi|_n$ cannot contain an index pair (k,l) with $a_{kl} = 0$, since otherwise D_π would be zero. Thus $\pi|_n \in P(A)$ and Claim 1 is proved.

To prove Claim 2, fix $\sigma \in P(A)$. We will in fact prove the stronger property that $\sum_{\pi:\pi|_n=\sigma} D_\pi = 1$. Consider a permutation $\pi \in S_{2n}$ with $\pi|_n = \sigma$. By the argument above, each element $h_{i,\sigma i}$ corresponds to two factors in D_π: either $b_{i,\sigma i}$ and $\overline{b}_{i,\sigma i}$ or $c_{i,\sigma i}$ and $-\overline{c}_{i,\sigma i}$. Thus the set $\{\pi : \pi|_n = \sigma\}$ is in 1-1 correspondence with the subsets of $[n]$, where the subset specifies those i which contribute factors $b_{i,\sigma i}$. Recall that for each i, either $b_{i,\sigma i} \in \mathbb{C}_4$ and $c_{i,\sigma_i} = 0$ or vice versa. Hence $D_\pi = 0$ for all but one of these permutations π, namely the permutation $\hat{\pi}$ corresponding to the subset $N = \{i : b_{i,\sigma i} \neq 0\}$. For this permutation, we have $D_{\hat{\pi}} = \mathrm{sgn}(\hat{\pi})(-1)^{n-|N|}$. And an easy induction on $n - |N|$ establishes that $\mathrm{sgn}(\hat{\pi}) = (-1)^{n-|N|}$, from which Claim 2 follows.

This concludes the proof of the first part of Theorem 4.2. □

4.3 Analysis of second moment

To prove the second moment claim in Theorem 4.2, we will proceed in similar fashion to the previous section. In particular, the main step once again is to express the second moment of the estimator for a general matrix A in terms of that for the 2×2 all-1's matrix A_1:

Theorem 4.3 *Let $E_1 = \mathrm{E}[Y_{A_1}^2]$. Then for any $n \times n$ $(0,1)$ matrix A, we have*

$$\mathrm{E}[Y_A^2] = \sum_{G \in \mathcal{G}(A)} (E_1 - 2)^{c(G)}.$$

The proof of Theorem 4.3 can be found in Appendix C. It is similar in overall structure to our second-moment analysis for general CL_m in section 3, but both simpler because of the fixed dimension $m = 3$ and more complex because of the Dieudonné determinant. This latter complication is handled with similar technology to the expectation analysis in the proof we have just given.

In light of Theorem 4.3, it remains only to compute E_1. Let $H = (h_{ij})$ be the random quaternion matrix computed by the algorithm when run on A_1. Following the progress of the algorithm Gauss shows that

$$Y_{A_1} = |\mathrm{Gauss}(H)|^2 = |h_{11}h_{22} - h_{11}h_{21}\overline{h}_{11}h_{12}|^2.$$

Thus Y_{A_1} has the same distribution as $|h-g|^2$, where h, g are chosen independently and u.a.r. from $\mathbb{H}_8$. An easy hand calculation then shows that Y_{A_1} takes the values 0 and 4 each with probability $\frac{1}{8}$, and the value 2 with probability $\frac{3}{4}$. Thus $E_1 = 5$, so from Theorem 4.3 we get $\mathrm{E}[Y_A^2] = \sum_{G \in \mathcal{G}(A)} 3^{c(G)}$ for an arbitrary A. Hence the critical ratio is bounded above by

$$\frac{\mathrm{E}[Y_A^2]}{\mathrm{E}[Y_A]^2} = \frac{\sum_{G \in \mathcal{G}(A)} 3^{c(G)}}{\sum_{G \in \mathcal{G}(A)} 2^{c(G)}} \leq \max_{G \in \mathcal{G}(A)} \left(\frac{3}{2}\right)^{c(G)} \leq \left(\frac{3}{2}\right)^{n/2}.$$

This completes the proof of the second claim in Theorem 4.2, and the analysis of our modified quaternion estimator. □

4.4 Beyond the quaternions

The question of whether the performance of our higher-dimensional Clifford algebra estimators X_A can also be achieved in polynomial time remains an intriguing open problem. Of course, a positive resolution would, in light of Theorem A, imply a fully-polynomial randomized approximation scheme for the permanent completely different from that of [10]. We have developed an efficient version of the estimator in the next algebra, CL_4, which according to large-scale experiments has the same performance as the corresponding X_A. This would actually overcome another major obstacle, as CL_4 is the first algebra that is not a division algebra (i.e., not all elements have inverses).[‖] However, so far we have not been able to go beyond this. We note that CL_m for any m has a representation as $k \times k$ matrices over $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$ (or direct sums of these), so one can always define an analog of the "reduced norm" we used for the quaternions. It is also easy to see that the resulting estimator is unbiased, but experiments indicate that the second moment is much larger than that of the corresponding X_A.

Acknowledgements

We would like to thank Sasha Barvinok, Amir Jafari, Laci Lovász, Mike Luby, Bernd Sturmfels and especially T-Y. Lam for helpful discussions.

References

[1] E. Artin, *Geometric Algebra*, Wiley Interscience, New York, 1988.

[2] H. Aslaksen, Quaternionic Determinants, *Mathematical Intelligencer* **18** (1996), 57–65.

[3] A. Barvinok, Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor, *Random Structures and Algorithms* **14** (1999), 29–61.

[4] A. Barvinok, "New permanent estimators via non-commutative determinants," Preprint, July 2000.

[5] A.Z. Broder, 'How hard is it to marry at random? (On the approximation of the permanent)," *Proceedings of the 18th Annual ACM Symposium on Theory of Computing* (STOC), ACM Press, 1986, 50–58. Erratum in *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, 1988, p. 551.

[6] C. Godsil and I. Gutman, "On the matching polynomial of a graph," *Algebraic Methods in Graph Theory*, 1981, pp. 241–249.

[7] A. Frieze and M. Jerrum, "An analysis of a Monte Carlo algorithm for approximating the permanent," *Combinatorica* **15** (1995), pp. 67–83.

[8] I. Gelfand and V. Retakh, "A theory of noncommutative determinants and characteristic functions of graphs," *Functional Analysis and Its Applications* **26** (1992), no. 4, pp. 1–20.

[9] M. Jerrum and A. Sinclair, "Approximating the permanent," *SIAM Journal on Computing* **18** (1989), 1149–1178.

[10] M. Jerrum, A. Sinclair and E. Vigoda, "A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries," *Proceedings of the 33rd ACM Symposium on Theory of Computing*, 2001, pp. 712–721.

[11] M. Jerrum and U. Vazirani, "A mildly exponential approximation algorithm for the permanent," *Algorithmica* **16** (1996), 392–401.

[‖] Note that the quaternions are known to have the highest dimension among all division algebras over the reals.

[12] N. Karmarkar, R. Karp, R. Lipton, L. Lovász, and M. Luby, "A Monte-Carlo algorithm for estimating the permanent," *SIAM Journal on Computing* **22** (1993), pp. 284–293.

[13] C. Kenyon, D. Randall, and A. Sinclair, "Approximating the number of dimer coverings of a lattice," *Journal of Statistical Physics* **83** (1996), pp. 637–659.

[14] T.-Y. Lam, *The algebraic theory of quadratic forms*, Benjamin/Addison-Wesley, Reading, MA, 1973. (Reprinted with revisions, 1980.)

[15] N. Linial, A. Samorodnitsky and A. Wigderson, "A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents," *Combinatorica* **20** (2000), 545–568.

[16] H. Minc, *Permanents*, Encyclopedia of Mathematics and its Applications **6** Addison-Wesley Publishing Company, 1982.

[17] N. Nisan, "Lower bounds for non-commutative computation," *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, 1991, pp. 410–418.

[18] L.E. Rasmussen, "Approximating the permanent: a simple approach," *Random Structures and Algorithms* **5** (1994), pp. 349–361.

[19] L.E. Rasmussen, "On approximating the permanent and other #P-complete problems," PhD Thesis, Computer Science Division, UC Berkeley, 1998.

[20] L.G. Valiant, "The complexity of computing the permanent," *Theoretical Computer Science* **8** (1979), 189–201.

[21] B.L. van der Waerden, *Algebra* Vol. 2, Frederick Ungar Publishing Co., New York, 1970.

Appendix

A The Clifford algebra CL_4

The Clifford algebra CL_4 has eight basis elements: $\{1, u_{12}, u_{23}, u_{13}, u_{1234}, u_{34}, u_{14}, u_{24}\}$. The only self-conjugate basis elements are 1 and u_{1234}. Note that these are also the only two elements that commute with all others. The complete multiplication table is as follows:

$$\begin{bmatrix} 1 & u_{12} & u_{23} & u_{13} & u_{1234} & u_{34} & u_{14} & u_{24} \\ u_{12} & -1 & u_{13} & -u_{23} & -u_{34} & u_{1234} & -u_{24} & u_{14} \\ u_{23} & -u_{13} & -1 & u_{12} & -u_{14} & u_{24} & u_{1234} & -u_{34} \\ u_{13} & u_{23} & -u_{12} & -1 & u_{24} & u_{14} & -u_{34} & -u_{1234} \\ u_{1234} & -u_{34} & -u_{14} & u_{24} & 1 & -u_{12} & -u_{23} & u_{13} \\ u_{34} & u_{1234} & -u_{24} & -u_{14} & -u_{12} & -1 & u_{13} & u_{23} \\ u_{14} & u_{24} & u_{1234} & u_{34} & -u_{23} & -u_{13} & -1 & -u_{12} \\ u_{24} & -u_{14} & u_{34} & -u_{1234} & u_{13} & -u_{23} & u_{12} & -1 \end{bmatrix}$$

If $h = c_1 + c_2u_{12} + c_3u_{23} + c_4u_{13} + c_5u_{1234} + c_6u_{34} + c_7u_{14} + c_8u_{24}$ then its conjugate $\overline{h}$ is defined as

$$\overline{h} = c_1 - c_2u_{12} - c_3u_{23} - c_4u_{13} + c_5u_{1234} - c_6u_{34} - c_7u_{14} - c_8u_{24}.$$

Note that

$$h\overline{h} = c_1^2 + c_2^2 + c_3^2 + c_4^2 + c_5^2 + c_6^2 + c_7^2 + c_8^2 + 2(c_1c_5 - c_2c_6 - c_3c_7 + c_4c_8)u_{1234},$$

which is not real. The norm-square is defined by $|h|^2 = \sum_{i=1}^{8} c_i^2$.

B Proofs from Section 3

In this section, we present proof details for Lemma 3.3, Lemma 3.5 and Theorem 3.6 that were omitted from the main text.

Proof of Lemma 3.3: Before proceeding with the case analysis of equation (2) outlined in the main text, we require some easy facts about self-conjugate subgroups.

Lemma B.1 *Let G be a subgroup of G_m, and let a be any element of G_m. Let H be the subset of G that commutes with a, i.e., $b \in H$ if and only if $ab = ba$. Then H is a subgroup of G, and furthermore, either $H = G$ or $|H| = |G|/2$.*

Proof: That H is a subgroup of G is immediate. Suppose that $H \neq G$, and let $G - H = \{b_1, \ldots b_r\}$ be those elements of G that do not commute with a. Now notice any product b_ib_j does belong to H. Thus the elements b_1b_i are all distinct and belong to H, so H is at least as large as $G - H$. □

Lemma B.2 *In the situation of the previous lemma, with $|H| = |G|/2$, let g be an element of G but not H. Then $gH = G - H$.*

Lemma B.3 [Expansion] *Let G be a self-conjugate subgroup of G_m, and let c be a self-conjugate element of G_m that commutes with G but is not in $G \cup -G$. Then $G \cup cG$ is also a self-conjugate subgroup of G_m and has twice the size of G.*

Lemma B.4 **[Conjugation]** *Let G be an arbitrary subgroup of G_m. For any $a \in G_m$, $a(\sum_{\alpha\in G} u_\alpha)\overline{a}$ can be written as $\sum_{\alpha\in G'} u_\alpha$, where $G' = aG\overline{a}$ is a conjugate subgroup of G. Hence if G is self-conjugate, then so is G'.*

We now proceed with the case analysis from the main part of the paper.

Case 1: $c \in G \cup -G$.

First, note that since G is self-conjugate and abelian then c is also self-conjugate and commutes with every element of G. Then expression (2) from the sketch proof of Lemma 3.3 in the main text becomes $2^k a(2\sum_{\alpha\in G} u_\alpha + (c+\overline{c})\sum_{\alpha\in G} u_\alpha)\overline{a} = 2^k a(2\sum_{\alpha\in G} u_\alpha + 2c\sum_{\alpha\in G} u_\alpha)\overline{a}$. Now if $c \in G$ then $cG = G$, so $c\sum_{\alpha\in G} u_\alpha = \sum_{\alpha\in G} u_\alpha$ and we get $2^{k+2}a(\sum_{\alpha\in G} u_\alpha)\overline{a}$. We then apply the conjugation lemma. Similarly, if $c \in -G$, then $cG = -G$ and we end up with 0.

Case 2: $c \notin G \cup -G$.

case 2a: c commutes with G and $c = \overline{c}$. As in case 1, (2) becomes $2^k a(2\sum_{\alpha\in G} u_\alpha + 2c\sum_{\alpha\in G} u_\alpha)\overline{a}$. Since c is self-conjugate and commutes with G, from the expansion lemma we can write this as $2^{k+1}a(\sum_{\alpha\in G'} u_\alpha)\overline{a}$ where $G' = G \cup cG$ is also self-conjugate. We then apply the conjugation lemma.

case 2b: All other cases. It is convenient to further divide this case as follows:

(i) c commutes with G but $c = -\overline{c}$. Here (2) becomes $2^{k+1}a(\sum_{\alpha\in G} u_\alpha)\overline{a}$, and the conjugation lemma finishes this case.

(ii) c does not commute with (all of) G. By Lemma B.1, the elements of G that commute with c form a subgroup $H \subset G$ with $|H| = |G|/2$. Thus the first two terms in the parentheses in (2) become $\sum_{\alpha\in G} u_\alpha + c(\sum_{\alpha\in G} u_\alpha)\overline{c} = 2\sum_{\alpha\in H} u_\alpha$. We now analyze the last two terms.

First suppose that $c = \overline{c}$. Here the last two terms become $c(\sum_{\alpha\in G} u_\alpha) + (\sum_{\alpha\in G} u_\alpha)\overline{c} = 2c\sum_{\alpha\in H} u_\alpha$ so, upon combining with the first two terms, (2) becomes $2^{k+1}a(\sum_{\alpha\in H} u_\alpha + c\sum_{\alpha\in H} u_\alpha)\overline{a}$. Now H must be self-conjugate since it is a subgroup of G; and c is self-conjugate, commutes with H, and does not belong to $H \cup -H$. Thus the expansion lemma allows us to rewrite $\sum_{\alpha\in H} u_\alpha + c\sum_{\alpha\in H} u_\alpha$ as $\sum_{\alpha\in G'} u_\alpha$, where $G' = H \cup cH$ is self-conjugate and $|G'| = 2|H| = |G|$. We then apply the conjugation lemma and we are done.

Now suppose that $c = -\overline{c}$. Here $c(\sum_{\alpha\in G} u_\alpha) + (\sum_{\alpha\in G} u_\alpha)\overline{c} = 2c\sum_{\alpha\in G-H} u_\alpha$, so upon combining with the first two terms (2) becomes $2^{k+1}a(\sum_{\alpha\in H} u_\alpha + c\sum_{\alpha\in G-H} u_\alpha)\overline{a}$. Recall from Lemma B.2 that $G - H = gH$ for some $g \in G$ not in H. Thus we can rewrite $\sum_{\alpha\in H} u_\alpha + c\sum_{\alpha\in G-H} u_\alpha$ as $\sum_{\alpha\in H} u_\alpha + cg\sum_{\alpha\in H} u_\alpha$. Since cg is self-conjugate ($(cg)^2 = cgcg = -ccgg = -c^2g^2 = -(-1)(1) = 1$) and commutes with H (both c and g commute with H), we can again apply the expansion lemma followed by the conjugation lemma.

This completes the case analysis and hence the proof of Lemma 3.3. □

In preparation for the proofs of Lemma 3.5 and Theorem 3.6, we develop some more detailed information about the subgroups arising in the proof of Lemma 3.3. For any particular G_m, we partition the set of all self-conjugate subgroups of G_m into equivalence classes such that $G \sim G'$ iff the respective probabilities (over the choice of a random signed basis element c) that G falls into cases 1,2a, and 2b are the same as those for G'. While it may appear *a priori* that we could have a proliferation of equivalence classes for each size of subgroup, this turns out not to be the case for most values of m:

Lemma B.5 *If* $m \neq 0 \bmod 4$ *then all subgroups of a given size* 2^i *that arise from the block diagonal estimator are equivalent.*

Proof of Lemma B.5: We do this by induction on the size 2^i of the subgroups. For the base case $i = 0$, we have only the trivial subgroup $\{1\}$, so the statement is vacuously true. Now assume $i > 0$. A subgroup G of size 2^i can arise either from expansion (as in case 2a) or from transmutation (as in the other cases). It therefore suffices to show the following two claims:

Claim 1: If subgroups G and G' of size 2^i are both formed by expansion, then G and G' are equivalent

Claim 2: If G transmutes to another group G'' of the same size, then G and G'' are equivalent.

Proof of Claim 1: We require the following lemmas:

Lemma B.6 *Suppose a subgroup H of size 2^{i-1} expands to a subgroup G of size 2^i. Then $|Z(H)| = 2|Z(G)|$, where $Z(H)$ and $Z(G)$ are the centralizers** of H and G in G_m.*

Proof: The proof appeals to a linear algebra description of G_m. Consider $\mathbb{F}_2^m$, the m-dimensional vector space over $\mathbb{F}_2$. We identify a basis element a of G_m with the vector v whose ℓth component is 1 if and only if ℓ appears among the subscripts of a. Note that given two basis elements a and b their product (up to sign) is described by the sum of their corresponding vectors $v + w$. Furthermore, if we define the dot product in the usual way, then a and b commute if and only if $v \cdot w = 0$.

A self-conjugate subgroup K of size 2^j can then be represented (up to sign) as a j-dimensional subspace W_K of $\mathbb{F}_2^m$. K is generated by exactly j elements, and a basis for the subspace W_K consists of the vectors corresponding to these generators. These vectors are distinct because of our restriction that if $a \in K$ then $a \notin -K$. The vectors must be linearly independent since a linear dependence would imply that one generator is a product of the others up to sign, which is impossible.

We define V_K to be the subspace spanned by W_K and $\overline{1}$, the all-ones vector. The basis elements that commute with K are then exactly those represented by the orthogonal subspace $V_K^\perp$. Any $v \in V_K^\perp$ must have an even number of subscripts since $v \cdot \overline{1} = 0$, and must commute with K since $v \cdot w = 0$ for all $w \in W_K$. It is a well-known fact that $\dim V_K + \dim V_K^\perp = \dim \mathbb{F}_2^m = m$. There are thus $2^{\dim V_K^\perp}$ unsigned basis elements in $Z(K)$, so $|Z(K)| = 2^{\dim V_K^\perp + 1}$.

Finally, observe that if H expands to G, then $\dim V_H = \dim V_G - 1$. This is true since G requires one more generator than H, and further, none of these generators can be $\overline{1}$ since $m \neq 0 \bmod 4$. Thus $|Z(H)| = 2|Z(G)|$. □

Lemma B.7 *When H expands to G as above, exactly half of the elements in $Z(H) - Z(G)$ are self-conjugate.*

Proof: Recall from case 2a of the proof of Lemma 3.3 that G can be described as $a(H \cup cH)\overline{a}$ for some self-conjugate c that commutes with H. Since conjugation by a changes only the signs of $H \cup cH$ we see that $Z(G) = Z(H \cup cH)$.

Note that $H \subseteq Z(H)$, and consider the set of cosets of H in $Z(H)$. Note that the elements of a coset are either all self-conjugate or all not self-conjugate ($(dh)^2 = dhdh = ddhh = d^2$). Also, the elements of a coset either all commute with c (and hence with G) or all do not commute with c ($dhc = cdh \Leftrightarrow dch = cdh \Leftrightarrow dc = cd$).

Now consider a coset dH in $Z(H) - Z(G)$ that is not self-conjugate. Then cdH is also in $Z(H) - Z(G)$ but is self-conjugate. Similarly, if dH in $Z(H) - Z(G)$ is self-conjugate, then cdH is not. Thus we have a bijection between self-conjugate and non-self-conjugate cosets in $Z(H) - Z(G)$, proving the lemma. □

To finish the proof of Claim 1, consider two subgroups G and G', both of size 2^i. Suppose that G expanded from H and G' from H'. By induction H and H' are equivalent. Thus $|Z(H)| = |Z(H')|$, and from Lemma B.6, we also have $|Z(G)| = |Z(G')|$. Since $|G| = |G'|$, the probabilities of case 1 are the same. From the equivalence of H and H' and Lemma B.6, we also have that the numbers

**Recall that the *centralizer* of a subgroup G in G_m is the set of elements of G_m that commute with all of G.

of elements of G_m that commute with G and G' are the same. Furthermore, from Lemma B.7 we have that the numbers of these that are self-conjugate are also the same. Hence the probabilities for case 2A are also equal, and therefore those for case 2B must be equal as well. □

Proof of Claim 2: Suppose a subgroup G transmutes to G'', as in cases 1 and 2B. We will show that G and G'' are equivalent. In fact, we will show the stronger result that there is some automorphism ϕ of G_m that carries G to G''. Recall that the analysis of case 2B in the proof of Lemma 3.3 included two subcases. In subcase (i), the new group G'' is of the form $aG\overline{a}$, and the automorphism is simply conjugation by a. The same argument applies in case 1.

Subcase (ii) of case 2B is slightly more involved. Recall that here we begin with G and a basis element c that commutes with only half of G. Then G can be described as $H \cup gH$, where H is the subgroup of G that commutes with c and $g \in G$ does not commute with c. If $c = \overline{c}$, we end up with $G'' = a(H \cup cH)\overline{a}$, and if $c = -\overline{c}$, we end up with $G'' = a(H \cup cgH)\overline{a}$. We claim the following, which is routinely verified:

Lemma B.8 *Let $d, g \in G_m$ be self-conjugate such that d does not commute with g. We say that an element $b \in G_m$ is* even *with respect to d and g if it commutes with neither or both of d and g, and* odd *otherwise. Then the map $\phi : G_m \to G_m$, defined as $\phi(b) = b$ for b even and $\phi(b) = dgb$ for b odd, is an automorphism of G_m.*

If $c = \overline{c}$, we apply this lemma with $d = c$ and see that $\phi(H \cup gH) = H \cup cH$, and if $c = -\overline{c}$ we set $d = cg$ and see that $\phi(H \cup gH) = H \cup cgH$. We then apply conjugation by a to obtain G''.

Once we have an automorphism, our claim follows immediately: if a basis element c falls into a certain case with respect to G, then $\phi(c)$ falls into the same case with respect to G''. □

As we are now done with Claims 1 and 2, this concludes the proof of Lemma B.5. □

The above analysis also allows us to precisely determine the size of the largest possible subgroup for all values of $m \neq 0 \bmod 4$. In particular we have the following:

Proposition B.9 *Let $m = 4q + 2$. Then the largest possible self-conjugate subgroup has size 2^{2q}.*

Proof: We prove this by determining how much a self-conjugate subgroup can expand, starting from the trivial subgroup $\{1\}$. There is no loss of generality here since any self-conjugate subgroup G can be seen as the result of a sequence of expansions. (For example, starting from $\{1\}$, we can iteratively expand to G by adding at each step an element of G not in the current subgroup.)

When $m \equiv 2 \bmod 4$, the self-conjugate elements comprise exactly half of the 2^{m-1} unsigned basis elements. (To see this, recall that the unsigned basis elements correspond to even cardinality subsets of $\{1, \ldots, m\}$; the self-conjugate elements correspond to those subsets of cardinality $4k+2$.) Thus when the subgroup has size $2^0 = 1$, its centralizer contains 2^{m-2} self-conjugate unsigned basis elements. At each expansion step, the subgroup size doubles and the number of self-conjuate unsigned basis elements in the centralizer halves until the two are the same; this occurs when the subgroup has size 2^t, where $t = \frac{m-2}{2} = 2q$. □

We assume from now on that $m = 4q + 2$. In this case, Proposition B.9 implies that the quantity $p = \max_G \frac{2|G|}{|G_m|}$ defined in Lemma 3.4 is exactly $\frac{1}{2^{2q+1}}$. This completes the deferred proof of Lemma 3.5.

Now for each possible size, 2^i, of self-conjugate subgroup, define $p_{i,k}(n)$ to be the probability that $\det(B_n)\overline{\det(B_n)}$ is of the form $2^k \sum_{\alpha \in G} u_\alpha$, where G is of size 2^i. (In light of Lemma B.5, this probability depends only on the size of G, not on its contents.) From our analysis in the proof of Lemma 3.3, we can write the following system of recurrence relations for the $p_{i,k}(n)$:

$$p_{i,k}(n) = \alpha_i p_{i,k-1}(n-1) + \beta_i p_{i,k-2}(n-1) + \gamma_i p_{i-1,k-1}(n-1), \tag{8}$$

where $p_{i,k}(0) = 1$ if $i = k = 0$ and $p_{i,k}(0) = 0$ for all other values of i, k. Here α_i is the probability of transmuting while doubling the coefficient (case 2B), β_i is the probability of transmuting while quadrupling (half of case 1), and γ_i is the probability of expanding (case 2A for the *smaller* group size 2^{i-1}). The values of the coefficients $\alpha_i, \beta_i, \gamma_i$ depend only on the sizes of G and its centralizer $Z(G)$, and the number of self-conjugate elements in $Z(G)$; thus they can be written down exactly using the technology developed in Lemmas B.6 and B.7.

We now write

$$P_i(n) = \sum_k 2^{2k} p_{i,k}(n)$$

so that the second moment of the estimator is

$$\mathrm{E}[X_{A_n}^2] = \sum_i P_i(n).$$

Theorem 3.6 is now an immediate consequence of the following two lemmas:

Lemma B.10 *For all i and all $n > 0$, $P_i(n) = \lambda_i P_i(n-1) + \mu_i P_{i-1}(n-1)$ for constants λ_i and μ_i (independent of n). Also $P_0(0) = 1$, $P_i(0) = 0$ otherwise.*

Lemma B.11 *The closed form of $\sum_i P_i(n)$ is a sum of two exponentials; i.e., $\sum_i P_i(n) = c_1 E_1^n + c_2 E_2^n$, where $E_1 = 4(1 + \frac{1}{2^{2q+1}})$, $E_2 = 4(1 - \frac{1}{2^{2q+1}})$, and c_1, c_2 are non-negative coefficients with $c_1 + c_2 = 1$.*

Proof of Lemma B.10: Observe the following:

$$\begin{aligned} P_i(n) &= \sum_k 2^{2k} p_{i,k}(n) \\ &= \sum_k 2^{2k} \Big(\alpha_i p_{i,k-1}(n-1) + \beta_i p_{i,k-2}(n-1) + \gamma_i p_{i-1,k-1}(n-1) \Big) \\ &= 4\alpha_i \sum_k 2^{2(k-1)} p_{i,k-1}(n-1) + 16\beta_i \sum_k 2^{2(k-2)} p_{i,k-2}(n-1) + 4\gamma_i \sum_k 2^{2(k-1)} p_{i-1,k-1} \\ &= (4\alpha_i + 16\beta_i) P_i(n-1) + 4\gamma_i P_{i-1}(n-1). \end{aligned}$$

Setting $\lambda_i = 4\alpha_i + 16\beta_i$ and $\mu_i = 4\gamma_i$ finishes the proof. $\square$

Proof of Lemma B.11: From Lemma B.10, we have a system of recurrences of the form $P_i(n) = \lambda_i P_i(n-1) + \mu_i P_{i-1}(n-1)$ for $i = 0, \ldots, t$, where $t = \frac{m-2}{2} = 2q$ (so the largest subgroup has size 2^t). Define $S(n) = \sum_i P_i(n)$, so that $S(n) = \mathrm{E}[X_{A_n}^2]$. Then $S(n) = \sum_{i=0}^{t} (\lambda_i + \mu_{i+1}) P_i(n-1)$, where we define $\mu_{t+1} = 0$. Now inspection of the proofs of Lemmas B.6 and B.7 allows us to write down exact values for $\alpha_i, \beta_i, \gamma_i$ in the recurrence (8), and hence for λ_i, μ_i. Bearing in mind the fact that $m = 4q + 2$, we get that $\mu_i = 4(\frac{1}{2^i} - \frac{1}{2^{m-i}})$ and $\lambda_i = 4(1 - \frac{1}{2^{i+1}} + \frac{1}{2^{m-i-2}})$. Hence we can write $S(n) = \sum_{i=0}^{t} 4(1 + \frac{2^i}{2^{m-1}}) P_{i-1}(n)$.

We now rewrite this as $S(n) = 4(1 + \frac{2^t}{2^{m-1}}) S(n-1) - R(n-1) = \lambda_t S(n-1) - R(n-1)$, where we define $R(n) = \sum_{i=0}^{t-1} 4(\frac{2^t - 2^i}{2^{m-1}}) P_i(n)$. Invoking the recurrence on the P_i, this can be written as $R(n) = \sum_{i=0}^{t-1} 4(\frac{2^t - 2^i}{2^{m-1}})(\lambda_i P_i(n-1) + \mu_i P_{i-1}(n-1))$. An elementary but tedious algebraic manipulation now shows that $R(n) = \lambda_{t-1} R(n-1)$, from which it follows that $R(n) = \lambda_{t-1}^n R(0)$. Here $R(0) = 4(\frac{2^t - 1}{2^{m-1}}) P_0(0) = \frac{2^t - 1}{2^{2t-1}}$.

Finally, plugging this back into the expression for $S(n)$ yields $S(n) = \lambda_t S(n-1) - \lambda_{t-1}^n R(0)$. Using the base case $S(0) = P_0(0) = 1$, this can be solved to give $S(n) = (1 - \frac{R(0)}{\lambda_t - \lambda_{t-1}}) \lambda_t^n + \frac{R(0)}{\lambda_t - \lambda_{t-1}} \lambda_{t-1}^n$.

Noting that $\lambda_t = E_1$ and $\lambda_{t-1} = E_2$, and that both coefficients are non-negative, we have proved the lemma. $\square$

C Proofs from Section 4

In this section, we supply the proof of Theorem 4.3 which was omitted from the main text.

Proof of Theorem 4.3: We require two straightforward preliminary observations about the behavior of the second moment on matrices whose graphs are the union of just two permutations.

Proposition C.1 *Let $G = G(\pi_1, \pi_2)$ for $\pi_1, \pi_2 \in S_n$ be any graph such that $c(G) = 1$. Then* $\mathrm{E}[Y^2_{A(G)}] = E_1$.

Proof: Recall that $|\mathrm{Gauss}(H)|^2 = \det(D_H)$ is invariant under elementary row and column operations on H. Thus we may assume that A has the form

$$A = \begin{pmatrix} I & 0 \\ 0 & A' \end{pmatrix}, \qquad \text{where} \qquad A' = \begin{pmatrix} 1 & & & & 1 \\ 1 & 1 & & \text{\Large 0} & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ \text{\Large 0} & & & 1 & 1 \end{pmatrix}.$$

Let H be the random quaternion matrix computed by the algorithm, and write $H = \begin{pmatrix} H_I & 0 \\ 0 & H' \end{pmatrix}$. Then clearly $\det(D_H) = \det(D_{H_I})\det(D_{H'})$, and it is easy to check that $\det(D_{H_I}) = 1$ always. Thus we need only prove the Proposition for A having the same bidiagonal form as A'. But since A_1 is the special case $n = 2$ of this form, it suffices to prove that $\mathrm{E}[Y_A^2]$ does not in this case depend on the size of A. The quaternion matrix H corresponding to such an A has the form

$$H = \begin{pmatrix} h_1 & & & & g_1 \\ g_2 & h_2 & & \text{\Large 0} & \\ & g_3 & h_3 & & \\ & & \ddots & \ddots & \\ \text{\Large 0} & & & g_n & h_n \end{pmatrix},$$

where the h_i and g_i are independent random elements of $\mathbb{H}_8$. When we run Gaussian elimination on this matrix, we get

$$\mathrm{Gauss}(H) = h_1 h_2 \cdots h_n + (-1)^{n+1} h_1 \cdots h_{n-1} g_n \overline{h}_{n-2} g_{n-2} \cdots g_2 \overline{h}_1 g_1.$$

But since h_n only appears as a factor in the first term, and g_n only in the second term, we see that $\mathrm{Gauss}(H)$ has the same distribution as $h + g$, where h and g are independent and uniformly distributed over $\mathbb{H}_8$. Thus $\mathrm{E}[Y_A^2]$ does not depend on n and we are done. $\square$

Proposition C.2 *Let $G = G(\pi_1, \pi_2)$ for $\pi_1, \pi_2 \in S_n$ be arbitrary. Then* $\mathrm{E}[Y^2_{A(G)}] = E_1^{c(G)}$.

Proof: As in the previous proof, by applying row and column operations we may assume that A has the form

$$A = \begin{pmatrix} A_1 & & & \text{\Large 0} \\ & A_2 & & \\ & & \ddots & \\ \text{\Large 0} & & & A_{c(G)} \end{pmatrix},$$

where each A_l is of the form dealt with in Proposition C.1. Let H be the quaternion matrix computed by the algorithm when run on A. Then H has the form

$$H = \begin{pmatrix} H_1 & & & \text{\Large 0} \\ & H_2 & & \\ & & \ddots & \\ \text{\Large 0} & & & H_{c(G)} \end{pmatrix},$$

where the matrices H_l are independent and distributed as if they had been computed by the algorithm run on the matrices A_l. Now because the norm-square function for the quaternions is a homomorphism (i.e., $|h_1h_2|^2 = |h_1|^2|h_2|^2$), we have

$$\det(D_H) = \prod_{l=1}^{c(G)} |\det(H_l)|^2,$$

which implies

$$\mathrm{E}[Y_A^2] = \prod_{l=1}^{c(G)} \mathrm{E}[Y_{A_l}^2].$$

The result now follows from Proposition C.1. □

We now proceed with the analysis of the second moment for a general matrix A. As in the analysis of the expectation, let $D = (d_{ij}) \in \mathbb{C}^{2n\times 2n}$ be the "reduced norm" matrix computed by the algorithm. Then we have

$$\mathrm{E}[X_A^2] = \mathrm{E}\left[\sum_{\pi,\pi'\in S_{2n}} \mathrm{sgn}(\pi\pi') \prod_{i=1}^{2n} d_{i,\pi i} d_{i,\pi' i}\right] \stackrel{\mathrm{def}}{=} \sum_{\pi,\pi'\in S_{2n}} \mathrm{E}[D_{\pi\pi'}]. \tag{9}$$

Again each entry d_{ij} of D depends on exactly one entry of H, namely $h_{\lceil \frac{i}{2}\rceil,\lceil \frac{j}{2}\rceil}$. Let $\pi,\pi' \in S_{2n}$. Then the factors $d_{i,\pi i}$ and $d_{i,\pi' i}$ of $D_{\pi\pi'}$ depend on a set $\pi\pi'|_n$ of between n and $4n$ entries of H, where $\pi\pi'|_n = \pi|_n \cup \pi'|_n$. From (9), Theorem 4.3 is an immediate consequence of the following two claims:

Claim 1: Let $\pi,\pi' \in S_{2n}$. Then $\mathrm{E}[D_{\pi\pi'}] \neq 0 \Rightarrow \pi\pi'|_n \in \mathcal{G}(A)$.

Claim 2: Let $G \in \mathcal{G}(A)$. Then $\mathrm{E}[D_G] \stackrel{\mathrm{def}}{=} \sum_{\pi,\pi':\pi\pi'|_n=G} \mathrm{E}[D_{\pi\pi'}] = (E_1-2)^{c(G)}$.

To prove Claim 1, fix $\pi,\pi' \in S_{2n}$ with $\mathrm{E}[D_{\pi\pi'}] \neq 0$. Clearly $\pi\pi'|_n$ corresponds to some bipartite graph $\mathcal{B}$ on $n+n$ vertices. To see that $\mathcal{B}$ must be a graph formed by two permutations in S_n, fix some odd $i = 2k-1 \in [2n]$ and consider the corresponding four factors, $d_{i,\pi i}$, $d_{i,\pi' i}$, $d_{i+1,\pi(i+1)}$, and $d_{i+1,\pi'(i+1)}$. Suppose $\pi i = 2l-1$ is odd (the case when it is even is similar). Thus $d_{i,\pi i} = b_{kl}$. Since the four factors under consideration are independent of all other factors in $D_{\pi\pi'}$, it must be the case that $\overline{b}_{kl}$ is also one of the four. In particular, either $\pi(i+1)$ or $\pi'(i+1)$ must equal $\pi i+1$ (so that either $d_{i+1,\pi(i+1)}$ or $d_{i+1,\pi'(i+1)}$ equals $\overline{b}_{kl}$). We assume $\pi'(i+1) = \pi i+1$; the other case is handled similarly. There are two cases to consider:

Case (i): $\pi' i = \pi i$, so that b_{kl} appears twice among the four factors. But then $\overline{b}_{kl}$ must also appear twice, so $\pi(i+1) = \pi i+1$. But then (k,l) is the only index pair in $\pi\pi'|_n$ of an entry in the kth row and lth column of H. Equivalently, $\mathcal{B}$ has an isolated edge from the kth vertex on the left to the lth vertex on the right.

Case (ii): $\pi' i \neq \pi i$. Observe that neither c_{kl} nor $-\overline{c}_{kl}$ can be a factor of $D_{\pi\pi'}$ since either $b_{kl} = 0$ or $c_{kl} = 0$. Thus $\lceil \frac{\pi' i}{2} \rceil \neq l$. So suppose $\pi' i = 2l'-1$ is odd with $l' \neq l$ (the even case is handled analogously). Then $\overline{b}_{kl'}$ must be the last of the four factors, i.e., $\pi(i+1) = 2l'$. But then (k,l) and (k,l') are the only index pairs in $\pi\pi'|_n$ of entries in the kth row of H. This corresponds to $\mathcal{B}$ having exactly two edges with incident on the kth vertex on the left. Now repeat the argument starting from the right vertices l and l' to see that they too have exactly two incident edges each. Repeat the argument until an even length cycle is closed.

The above two cases imply that each vertex on the left of $\mathcal{B}$ is either the endpoint of an isolated edge, or takes part in an even length cycle. We conclude that $\pi\pi'|_n$ is a graph formed by two permutations in S_{2n}. Clearly $\pi\pi'|_n$ cannot contain an index pair (k,l) with $a_{kl} = 0$. Claim 1 follows.

We now prove Claim 2 by induction on $c(G)$. Fix $G \in \mathcal{G}(A)$. For the base case, assume $c(G) = 0$. But then $G = \pi$ for some permutation $\pi \in P(A)$, and it is easy to see that $\mathrm{E}[D_G] = 1 = (E_1-2)^0$.

For the induction step, assume the claim holds true for all graphs G' with fewer cycles than G. Consider the matrix $A(G)$. It follows from Claim 1 that

$$\mathrm{E}[X_{A(G)}^2] = \sum_{G' \in \mathcal{G}(A(G))} \mathrm{E}[D_{G'}].$$

But now by Proposition C.2 we may write

$$E_1^{c(G)} = \sum_{G' \subseteq G} \mathrm{E}[D_{G'}],$$

and we may apply the induction hypothesis:

$$\begin{aligned} E_1^{c(G)} &= \mathrm{E}[D_G] + \sum_{k=1}^{c(G)} \tbinom{c(G)}{k} 2^k (E_1 - 2)^{c(G)-k} \\ &= \mathrm{E}[D_G] + E_1^{c(G)} - (E_1 - 2)^{c(G)}. \end{aligned}$$

Thus we have $\mathrm{E}[D_G] = (E_1 - 2)^{c(G)}$, which completes the proof of Claim 2 and of Theorem 4.3. $\square$