Generalized Kakeya sets for polynomial evaluation and faster computation of fermionants

Algorithmica
https://doi.org/10.1007/s00453-018-0513-7

Generalized Kakeya sets for polynomial evaluation and faster computation of fermionants

Andreas Björklund[1] · Petteri Kaski[2] · Ryan Williams[3]

Received: 31 December 2017 / Accepted: 30 August 2018

Abstract

We present two new data structures for computing values of an n-variate polynomial P of degree at most d over a finite field of q elements. Assuming that d divides $q-1$, our first data structure relies on $(d+1)^{n+2}$ tabulated values of P to produce the value of P at any of the q^n points using $O(nqd^2)$ arithmetic operations in the finite field. Assuming that s divides d and d/s divides $q-1$, our second data structure assumes that P satisfies a degree-separability condition and relies on $(d/s+1)^{n+s}$ tabulated values to produce the value of P at any point using $O(nq^s sq)$ arithmetic operations. Our data structures are based on generalizing upper-bound constructions due to Mockenhaupt and Tao (Duke Math J 121(1):35–74, 2004), Saraf and Sudan (Anal PDE 1(3):375–379, 2008) and Dvir (Incidence theorems and their applications, 2012. arXiv:1208.5073) for Kakeya sets in finite vector spaces from linear to higher-degree polynomial curves. As an application we show that the new data structures enable a faster algorithm for computing integer-valued *fermionants*, a family of self-reducible polynomial functions introduced by Chandrasekharan and Wiese (Partition functions of strongly correlated electron systems as fermionants, 2011. arXiv:1108.2461v1) that captures numerous fundamental algebraic and combinatorial functions such as the determinant, the permanent, the number of Hamiltonian cycles in a directed multigraph, as well as certain partition functions of strongly correlated electron systems in statistical physics. In particular, a corollary of our main theorem for fermionants is that the permanent of an $m \times m$ integer matrix with entries bounded in absolute value by a constant can be computed in time $2^{m-\Omega(\sqrt{m/\log\log m})}$, improving an earlier algorithm of Björklund (in: Proceedings of the 15th SWAT, vol 17, pp 1–11, 2016) that runs in time $2^{m-\Omega(\sqrt{m/\log m})}$.

✉ Petteri Kaski
petteri.kaski@aalto.fi

[1] Department of Computer Science, Lund University, Lund, Sweden

[2] Department of Computer Science, Aalto University, Helsinki, Finland

[3] Department of Electrical Engineering and Computer Science, CSAIL, MIT, Cambridge, MA, USA

Published online: 18 September 2018

Keywords Besicovitch set · Fermionant · Finite field · Finite vector space · Hamiltonian cycle · Homogeneous polynomial · Kakeya set · Permanent · Polynomial evaluation · Tabulation

1 Introduction

The protagonist of this paper is the following task. We want an efficient representation of an n-variate degree-d polynomial P over a finite field $\mathbb{F}_q$ of order q, that permits us to evaluate P on arbitrary points $a \in \mathbb{F}_q^n$. What kind of resource trade-offs can be achieved between space (for representing P) and query time (for computing $P(a)$ at a given a)?

The study of data structures that enable fast "polynomial evaluation" queries for multivariate polynomials was initiated by Kedlaya and Umans [12] for polynomials with bounded individual variable degrees, motivated by applications to fast polynomial factorization. Here we focus on the case when P has (total) degree d, in particular, when d is less than n.[1]

We seek data structures consisting of a set $K \subseteq \mathbb{F}_q^n$ and an associated list $((a, P(a)) : a \in K)$ of evaluations. There are two extremes for such designs. At one extreme, we can set $K = \mathbb{F}_q^n$, put all evaluations in a sorted array, and assuming constant-time random access, we achieve $O(n)$ query time. At the other extreme, to uniquely identify P we must tabulate $\Omega\binom{n+d}{d}$ points, as this is the dimension of the monomial basis. However, when K is this small, we are only aware of brute-force ($n^{O(d)}$-time) algorithms to evaluate the polynomial in any other point. Between these two extremes, we seek constructions for sets K that suffice for evaluating P at any point outside K in time that scales sub-exponentially in d. Our motivation is to accelerate the best known algorithms for canonical #P-hard problems (cf. Sect. 1.2).

1.1 Polynomial Evaluation Based on Generalized Kakeya Sets

Let $\mathbb{F}_q[x]$ be the ring of polynomials over indeterminates $x = (x_1, x_2, \ldots, x_n)$ with coefficients in the finite field $\mathbb{F}_q$. For fundamentals of finite fields, we refer to Lidl and Niederreiter [17]. Let us write $\mathrm{M}(q)$ for the time complexity of multiplication and division in $\mathbb{F}_q$. For example, $\mathrm{M}(q) = O\big((\log q)^{1+\epsilon}\big)$ holds for any constant $\epsilon > 0$; we refer to e.g. von zur Gathen and Gerhard [26] for sharper bounds. For fields of order $q = p^d$ for a prime p and an integer $d \geq 2$ we tacitly assume here and in what follows that an irreducible polynomial of degree d in $\mathbb{F}_p[x]$ is available to support the computations in $\mathbb{F}_q$ (cf. [26, §14.9] for construction algorithms for irreducible polynomials over $\mathbb{F}_p$).

Our first main theorem constructs an explicit set $K \subseteq \mathbb{F}_q^n$ of cardinality at most $(d+1)^{n+2}$ which allows for relatively quick evaluation of any degree-d P at all points in $\mathbb{F}_q^n$.

[1] In contrast, Kedlaya and Umans [12] focus on the case $n \leq d^{o(1)}$; cf. [12, Corollaries 4.3, 4.5, and 6.4]. *Notational caveat:* Kedlaya and Umans use "m" for the number of variables.

Theorem 1 *Let d divide $q-1$. There is a set $K \subseteq \mathbb{F}_q^n$ of size $|K| \le (d+1)^{n+2}$ along with functions $g_1, g_2, \ldots, g_{(q-1)(d+1)^2} : \mathbb{F}_q^n \to K$ and scalars $\gamma_1, \gamma_2, \ldots, \gamma_{(q-1)(d+1)^2} \in \mathbb{F}_q$ such that for every polynomial $P \in \mathbb{F}_q[x]$ of degree at most d and every vector $a \in \mathbb{F}_q^n$,*

$$P(a) = \sum_{j=1}^{(q-1)(d+1)^2} \gamma_j P(g_j(a))\,.$$

Moreover, there is an algorithm that in time $O(|K| nq\mathrm{M}(q))$ lists the elements of K, and there is an algorithm that in time $O\big(nqd^2\mathrm{M}(q)\big)$ computes the values $g_j(a) \in \mathbb{F}_q^n$ and $\gamma_j \in \mathbb{F}_q$ for all $j = 1, 2, \ldots, (q-1)(d+1)^2$ when given $a \in \mathbb{F}_q^n$ as input.

The size of K can be further reduced for polynomials P satisfying a certain (natural) restriction which holds for several well-studied polynomials. Suppose we partition the variable set $X = \{x_1, x_2, \ldots, x_n\}$ into

$$X = X_1 \cup X_2 \cup \cdots \cup X_d$$

such that

$$|X_1| = |X_2| = \cdots = |X_d| = \frac{n}{d}\,.$$

Let us say that a degree-d polynomial $P \in \mathbb{F}_q[x]$ is *degree-separable* relative to $X_1, X_2, \ldots, X_d$ if every monomial of P contains one variable from each X_i. Note a degree-separable P is in particular both multilinear and homogeneous of degree d. Degree-separability enables a trade-off between the size of K and the query time for evaluation:

Theorem 2 *Let s divide d and d/s divide $q-1$. There is a set $K \subseteq \mathbb{F}_q^n$ of size $|K| \le (d/s+1)^{n+s}$ along with $g_1, g_2, \ldots, g_{(q-1)^s} : \mathbb{F}_q^n \to K$ and $\gamma_1, \gamma_2, \ldots, \gamma_{(q-1)^s} \in \mathbb{F}_q$ such that for every degree-separable degree-d $P \in \mathbb{F}_q[x]$ relative to a fixed partition $X_1, X_2, \ldots, X_d$ and every vector $a \in \mathbb{F}_q^n$,*

$$P(a) = \sum_{j=1}^{(q-1)^s} \gamma_j P(g_j(a))\,.$$

Moreover, there is an algorithm that in time $O(|K| nq\mathrm{M}(q))$ lists the elements of K, and there is an algorithm that in time $O(n(q-1)^s sq\mathrm{M}(q))$ computes the values $g_j(a) \in \mathbb{F}_q^n$ and $\gamma_j \in \mathbb{F}_q$ for all $j = 1, 2, \ldots, (q-1)^s$ when given $a \in \mathbb{F}_q^n$ as input.

We need K to contain enough points that "interpolation" at all the other points is possible. One intuition for designing a small $K \subseteq \mathbb{F}_q^n$ for polynomial evaluation is that such a set must enable "localization" of any target polynomial inside the set. At one extreme, we may think of the simplest non-constant family of polynomials, namely *lines*. In Euclidean spaces, this line of thought leads to the study of dimensionality of sets that contain a unit line segment in every direction, or the *Kakeya problem*, which has been extensively studied since the 1920s and the seminal work of Besicovitch [2]. We refer to Wolff [28], Mockenhaupt and Tao [20], and Dvir [9,10] for surveys both in the continuous and finite settings. In what follows we focus on finite vector spaces.

Definition 1 A *Kakeya set* (or *Besicovitch set*) in a vector space of dimension n over $\mathbb{F}_q$ is a subset $K \subseteq \mathbb{F}_q^n$ together with a function $f : \mathbb{F}_q^n \to \mathbb{F}_q^n$ such that for every vector $a \in \mathbb{F}_q^n$ and every scalar $\tau \in \mathbb{F}_q$ it holds that

$$f(a) + \tau a \in K \,. \tag{1}$$

That is, a Kakeya set K has the property that for any possible direction of a line in $\mathbb{F}_q^n$ (that is, any nonzero vector $a \in \mathbb{F}_q^n$), the set K contains an entire line (through $f(a)$) with this direction. To support our objective of polynomial evaluations for higher-order curves than lines, an intuition is now to generalize (1) to polynomials of higher degree in the indeterminate τ. This is the methodological gist of our main contribution in this paper, which will be described further in Sect. 2.

As an illustrative application of our new data structures, we use Theorem 2 to derive a faster algorithm for computing fermionants, which are a family of self-reducible and degree-separable polynomials introduced by Chandrasekharan and Wiese [7] to generalize various fundamental polynomials. We start with a brief introduction to fermionants to motivate their study from a computational perspective.

1.2 Fermionants

We continue to work over $\mathbb{F}_q$. As usual, S_m is the symmetric group over $[m] = \{1, 2, \ldots, m\}$. We write $c(\sigma)$ for the number of cycles in a permutation $\sigma \in S_m$, where each fixed point of σ is counted as a cycle of length 1. Let $A = (a_{ij} : i, j \in [m])$ be an $m \times m$ matrix of indeterminates. The *fermionant* of A with (indeterminate) parameter t is the $(m^2 + 1)$-variable polynomial

$$\operatorname{fer}_t A = (-1)^m \sum_{\sigma \in S_m} (-t)^{c(\sigma)} \prod_{i=1}^{m} a_{i,\sigma(i)} \,. \tag{2}$$

The fermionant is multilinear and homogeneous of degree m with respect to the variables $\{a_{i,j}\}$, and of degree m with respect to t. Furthermore, note that with respect to $\{a_{i,j}\}$ the fermionant is degree-separable under the partition $\{\{a_{ij} : j \in [m]\} : i \in [m]\}$.

The fermionant captures several extensively studied algebraic and combinatorial functions, such as the determinant of a matrix

$$\det A = (-1)^m \sum_{\sigma \in S_m} (-1)^{c(\sigma)} \prod_{i=1}^{m} a_{i,\sigma(i)} \,,$$

the permanent of a matrix

$$\operatorname{per} A = \sum_{\sigma \in S_m} \prod_{i=1}^{m} a_{i,\sigma(i)} \,,$$

the generating function for directed Hamiltonian cycles

$$\operatorname{hc} A = \sum_{\substack{\sigma \in S_m \\ c(\sigma)=1}} \prod_{i=1}^{m} a_{i,\sigma(i)} \,,$$

as well as certain partition functions of strongly correlated electron systems in statistical physics (see Chandrasekharan and Wiese [7]). It is immediate that the aforementioned functions can be obtained as special cases of the fermionant via

$$\begin{aligned} \det A &= \operatorname{fer}_1 A \,, \\ \operatorname{per} A &= (-1)^m \operatorname{fer}_{-1} A \,, \\ \operatorname{hc} A &= (-1)^{m-1} \{t\} \operatorname{fer}_t A \,, \end{aligned}$$

where in the last equality we write $\{t^k\}P$ for the coefficient of t^k in the polynomial P.

The functions captured by the fermionant have received such substantial attention that is not possible to discuss the literature exhaustively here. For example, the permanent and the determinant are central to arithmetic circuit complexity [24] and geometric complexity theory [16]. Similarly, the numerous symmetries and self-reducibility properties of fermionants enable their use in e.g. interactive proof systems [5,18,27]. We restrict our present discussion of earlier work mostly to algorithms for the permanent.

Computing the permanent of a given $m \times m$ matrix appears to be an extremely hard problem. The restriction to $0/1$-matrices is equivalent to the problem of counting perfect matchings in a bipartite graph with the matrix equal to the graph's biadjacency matrix. The best known general algorithm is over 50 years old, given by Ryser [21] in 1963, and it uses $O(2^m m)$ arithmetic operations. Valiant [25] proved that the permanent for $\{0, 1\}$-matrix inputs is #P-hard, even if the number of ones per row is at most three. In the more general setting of fermionants, Mertens and Moore [19] showed that the fermionant fer_t is #P-hard for any $t > 2$ and $\oplus$P-hard for $t = 2$, even for the adjacency matrices of planar graphs. For the permanent, no less-than-2^m-sized arithmetic circuit is known despite substantial efforts (for example, it is a prominent open problem in the *Art of Computer Programming* [13]).

However, there are faster ways to compute the permanent if we allow random-access tabulation *along with* arithmetic operations. Most notably, there are modest speed-ups for $\{0, 1\}$-matrices over the integers. Bax and Franklin [1] gave an $2^{m-\Omega(m^{1/3}/\log m)}$ expected time algorithm. Björklund [3] presented a deterministic $2^{m-\Omega(\sqrt{m/\log q})}$ time algorithm over any finite field of order $q \geq m^2 + 1$, by exploiting the self-reducibility of the permanent. Applying the Chinese Remainder Theorem, he also obtains a $2^{m-\Omega(\sqrt{m/\log m})}$-time algorithm for integer matrices with entries whose absolute value is bounded from above by a constant. There are also faster algorithms for sparse matrices. Cygan and Pilipczuk [8] gave a $2^{m-\Omega(m/r)}$ time algorithm for matrices with at most r non-zero entries per row. Very recently, Björklund, Husfeldt, and Lyckberg [4] and Björklund, Kaski, and Koutis [6] show that if the result is bounded in absolute value by c^m for a constant $c > 1$, then there are $2^{m\left(1-1/c^{\Omega(1)}\right)}m^{O(1)}$-time algorithms for the permanent and the number of directed Hamiltonian cycles, respectively. Both algorithms work by computing the permanent and the number of directed Hamiltonian cycles modulo small primes. In particular, the algorithms over $\mathbb{F}_p$ run in time $2^{m\left(1-1/c^{\Omega(1)}\right)}m^{O(1)}$, faster than the algorithms of this paper for small p.

Our main technical result for fermionants is that, given mild technical conditions on the order of the field, we can obtain a faster algorithm over finite fields:

Theorem 3 *There is an algorithm that computes the coefficients of the fermionant as a polynomial in t, $\mathrm{fer}_t\, A \in \mathbb{F}_q[t]$ of a given matrix $A \in \mathbb{F}_q^{m\times m}$ in time $2^{m-\Omega(\sqrt{m/\log\log q})}\, O\,(\mathrm{M}(q))$, provided that $q-1$ has a divisor in the interval $(1.1\log q, 10\log q)$, $q \geq m^2+1$, and $m = \omega\big(\log^2 q \log\log q\big)$.*

The Chinese Remainder Theorem and a uniform variant of the Prime Number Theorem for arithmetic progressions yield the following corollary for integer-valued fermionants.

Corollary 1 *Let t be an integer with $|t|$ in $O(m)$ and let M be a constant. The fermionant $\mathrm{fer}_t\, A$ can be computed in time $2^{m-\Omega(\sqrt{m/\log\log m})}$, for all $m \times m$ matrices A with integer values in $[-M, M]$.*

As far as we know, our result gives the currently fastest algorithms to compute a bounded integer entry permanent and counting the number of Hamiltonian cycles in a directed graph. The idea behind Theorem 3 is to apply our polynomial evaluation results to a self-reduction for fermionants. Following Björklund's results for the permanent [3], we show how to compute a fermionant on an $m \times m$ matrix via $2^{m-k}m^{O(1)}$ calls to the fermionant on $k \times k$ matrices. Applying Theorem 2, we set k so that it is possible to evaluate the $k \times k$ fermionant polynomial over all points of K in $2^{0.999m}$ time. Once we know the polynomial on all points in K, we can then evaluate the fermionant on any $m \times m$ matrix in time about $2^{m-\Omega(k)}m^{O(1)}$. We show $k \approx \sqrt{m/\log\log q}$ suffices.

1.3 Organisation of the Paper

In Sect. 2, we present our generalization of Kakeya sets in finite vector spaces, together with explicit constructions. Next in Sect. 3 we prove our main evaluation theorems, Theorem 1 and Theorem 2. In Sect. 4 we use the self-reducibility of the fermionant to prove Theorem 3 and Corollary 1, showing how to compute fermionants faster.

2 Generalized Kakeya Sets in Finite Vector Spaces

We study the following generalization of Kakeya sets for lines (Definition 1) to higher-degree polynomial curves:

Definition 2 A *Kakeya set* of *degree* r in a vector space of dimension n over $\mathbb{F}_q$ consists of a set $K \subseteq \mathbb{F}_q^n$ together with functions $f_0, f_1, \ldots, f_{r-1} : \mathbb{F}_q^n \to \mathbb{F}_q^n$ such that for every vector $a \in \mathbb{F}_q^n$ and every scalar $\tau \in \mathbb{F}_q$ it holds that

$$F(a, \tau) = f_0(a) + f_1(a)\tau + f_2(a)\tau^2 + \ldots + f_{r-1}(a)\tau^{r-1} + a\tau^r \in K\,. \quad (3)$$

We say that a construction for Kakeya sets is *explicit* if

1. there is an algorithm that outputs K (given q, r, and n) in $O\big(|K| nr \mathrm{M}(q)\big)$ time, and
2. there is an algorithm that given $a \in \mathbb{F}_q^n$ outputs the values

$$f_0(a), f_1(a), \ldots, f_{r-1}(a) \in \mathbb{F}_q^n$$

in $O\big(nr\mathrm{M}(q)\big)$ time.

The following construction of sparse Kakeya sets of degree r generalizes the design of the best known Kakeya sets (cf. Mockenhaupt and Tao [20], Saraf and Sudan [22], Dvir [9, §2.4], Kopparty, Lev, Saraf, and Sudan [14], and Kyureghyan, Müller, and Wang [15]).

Lemma 1 *For every $r+1$ that divides $q-1$ there is an explicit Kakeya set $K \subseteq \mathbb{F}_q^n$ of degree r and size $|K| \le \left(\frac{q-1}{r+1} + 1\right)^{n+1}$.*

Proof We begin with three simple observations. First, since $r+1$ divides $q-1$, we have that $r+1$ has a multiplicative inverse in $\mathbb{F}_q$. Indeed, $q = p^a$ for a prime p and positive integer a. Notice that $r+1$ has a multiplicative inverse if and only if p does not divide $r+1$. By assumption we have $(r+1)Q = p^a - 1$ for an integer Q and thus $r+1 = pb$ for an integer b would lead to a contradiction $p(bQ - p^{a-1}) = 1$. Second, for all $\alpha, \tau \in \mathbb{F}_q$ from the Binomial Theorem we have

$$\left(\frac{\alpha}{r+1} + \tau\right)^{r+1} - \tau^{r+1} = \sum_{i=0}^{r-1} \binom{r+1}{i} \left(\frac{\alpha}{r+1}\right)^{r+1-i} \tau^i \; + \; \alpha\tau^r\,. \quad (4)$$

Third, since the multiplicative subgroup $\mathbb{F}_q^\times$ is cyclic of order $q-1$, the subgroup consisting of $(r+1)$th powers of elements of $\mathbb{F}_q^\times$ has size exactly $\frac{q-1}{r+1}$. Including the zero element, we observe that $|\{\beta^{r+1} : \beta \in \mathbb{F}_q\}| = \frac{q-1}{r+1} + 1$.

Let us now define $K \subseteq \mathbb{F}_q^n$ to consist of all vectors of the form

$$\left(\left(\frac{\alpha_1}{r+1} + \tau\right)^{r+1} - \tau^{r+1}, \left(\frac{\alpha_2}{r+1} + \tau\right)^{r+1} - \tau^{r+1}, \ldots, \left(\frac{\alpha_n}{r+1} + \tau\right)^{r+1} - \tau^{r+1}\right) \quad (5)$$

with $\alpha_1, \alpha_2, \ldots, \alpha_n, \tau \in \mathbb{F}_q$. It follows immediately from (5) and our third observation that

$$|K| \leq \left(\frac{q-1}{r+1} + 1\right)^{n+1} .$$

Furthermore, (4) and (5) imply that the generalized Kakeya property (3) holds when we define the functions $f_i : \mathbb{F}_q^n \to \mathbb{F}_q^n$ for all $i = 0, 1, \ldots, r-1$ and $a = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{F}_q^n$ by

$$f_i(a) = \left(\binom{r+1}{i}\left(\frac{\alpha_1}{r+1}\right)^{r+1-i}, \ldots, \binom{r+1}{i}\left(\frac{\alpha_n}{r+1}\right)^{r+1-i}\right). \tag{6}$$

It is immediate from the definitions (5) and (6) that the construction is explicit. □

3 Polynomial Evaluation

This section proves our two main theorems for polynomial evaluation. The key idea is Mellin-transform-like sieving (8) enabled by an elementary observation about sums over finite fields (7) below, which we then extend to an s-fold product form in (13).

Let us start with a homogeneous version of Theorem 1.

Lemma 2 *Let d divide $q-1$. There is a set $K \subseteq \mathbb{F}_q^n$ of size $|K| \leq (d+1)^{n+1}$ together with functions $g_1, g_2, \ldots, g_{q-1} : \mathbb{F}_q^n \to K$ and scalars $\gamma_1, \gamma_2, \ldots, \gamma_{q-1} \in \mathbb{F}_q$ such that for every homogeneous polynomial $P \in \mathbb{F}_q[x]$ of degree $h \leq d$ and every vector $a \in \mathbb{F}_q^n$,*

$$P(a) = \sum_{j=1}^{q-1} \gamma_j P(g_j(a)) .$$

Moreover, there is an algorithm that in time $O(|K| nq \mathrm{M}(q))$ lists the elements of K, and there is an algorithm that in time $O(nq\mathrm{M}(q))$ computes the values $g_j(a) \in \mathbb{F}_q^n$ and $\gamma_j \in \mathbb{F}_q$ for all $j = 1, 2, \ldots, (q-1)$ when given $a \in \mathbb{F}_q^n$ as input.

Proof Set $r = (q-1)/d - 1$, and note that $r+1$ divides $q-1$. Apply Lemma 1 to obtain K and the functions $f_0, f_1, \ldots, f_{r-1}$. Let $P \in \mathbb{F}_q[x]$ be a homogeneous polynomial of degree $h \leq d$ over the indeterminates $x = (x_1, x_2, \ldots, x_n)$, and let $a = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{F}_q^n$ be an assignment of values to the indeterminates. Our goal is to compute the value $P(a) \in \mathbb{F}_q$ using evaluations of P at K. Recalling the function $F(a, \tau)$ from (3), we will rely on values of the composition $P(F(a, \tau))$ for $\tau \in \mathbb{F}_q$ to obtain $P(a)$. Towards this end, we first observe that

$$\sum_{\tau \in \mathbb{F}_q^\times} \tau^e = \begin{cases} -1 & \text{if} \quad q-1 \text{ divides } e, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

To see this, let g be a generator of the multiplicative subgroup $\mathbb{F}_q^\times$. If $q-1$ divides e then $\tau^e = 1$ for all τ, and thus the sum is $|\mathbb{F}_q^\times| = q-1$ (modulo the characteristic). Otherwise, $g^e \neq 1$, and we have

$$\sum_{\tau \in \mathbb{F}_q^\times} \tau^e = \sum_{\tau \in \mathbb{F}_q^\times} (g\tau)^e = g^e \sum_{\tau \in \mathbb{F}_q^\times} \tau^e ,$$

so the sum must be 0.

Let $t = q-1-rh$ and observe that $t \geq 1$. We now claim that

$$P(a) = -\sum_{\tau \in \mathbb{F}_q^\times} \tau^t P(F(a,\tau)) . \tag{8}$$

By linearity, it suffices to consider the case when P is a single monomial $P = x_1^{h_1} x_2^{h_2} \cdots x_n^{h_n}$ of degree $h = h_1 + h_2 + \ldots + h_n \leq d$. Recalling (3) and (7), we observe that the right-hand side of (8) expands to

$$\begin{aligned}
&-\sum_{\tau \in \mathbb{F}_q^\times} \tau^t P(F(a,\tau)) \\
&\quad = -\sum_{\tau \in \mathbb{F}_q^\times} \tau^{q-1-rh} \left(\tau^{rh} \alpha_1^{h_1} \alpha_2^{h_2} \cdots \alpha_n^{h_n} + \tau^{rh-1}(\cdots) + \ldots + \tau^0(\cdots)\right) \\
&\quad = -\sum_{\tau \in \mathbb{F}_q^\times} \left(\tau^{q-1} \alpha_1^{h_1} \alpha_2^{h_2} \cdots \alpha_n^{h_n} + \tau^{q-2}(\cdots) + \ldots + \tau^{q-1-rh}(\cdots)\right) \\
&\quad = \alpha_1^{h_1} \alpha_2^{h_2} \cdots \alpha_n^{h_n} \\
&\quad = P(a) .
\end{aligned}$$

That is, by multiplying each term by τ^t, we ensure that all other terms appearing inside of $P(F(a,\tau))$ cancel, except for the desired term $\alpha_1^{h_1} \alpha_2^{h_2} \cdots \alpha_n^{h_n}$ which is the coefficient of τ^{rh}.

Now let $\beta_1, \beta_2, \ldots, \beta_{q-1}$ be an enumeration of the elements of $\mathbb{F}_q^\times$. For all $j = 1, 2, \ldots, q-1$, set $g_j(a) = F(a, \beta_j)$ and $\gamma_j = -\beta_j^t$. The first part of the lemma now follows from (8).

The running time bounds follow from the fact that the construction in Lemma 1 is explicit and $d < q$. □

3.1 Proof of Theorem 1

We are now ready to prove Theorem 1. Our strategy is to interpolate the homogeneous components of our given polynomial, then apply Lemma 2. Towards this end, let $P \in \mathbb{F}_q[x]$ have degree at most d and let $P = \sum_{h=0}^{d} P_h$ where $P_h \in \mathbb{F}_q[x]$ is either zero or homogeneous of degree h, for all $h = 0, 1, \ldots, d$. Let $\nu_0, \nu_1, \ldots, \nu_d$ be any

$d+1$ distinct elements of $\mathbb{F}_q$. Recalling the definition of K in (5), let $\hat{K} \subseteq \mathbb{F}_q^n$ be the set of all vectors of the form

$$\nu\left(\left(\frac{\alpha_1}{r+1}+\tau\right)^{r+1}-\tau^{r+1}, \ldots, \left(\frac{\alpha_n}{r+1}+\tau\right)^{r+1}-\tau^{r+1}\right) \tag{9}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n, \tau \in \mathbb{F}_q$, and $\nu \in \{\nu_0, \nu_1, \ldots, \nu_d\}$.

In particular, from (9) and (5) we have that $|\hat{K}| \leq (d+1)|K|$.

Assuming we have constant-time access to $P(a)$ for all $a \in \hat{K}$, we can access each P_h at $k \in K$ by univariate interpolation over the $d+1$ distinct values of ν, via the identity

$$P(\nu k) = \sum_{h=0}^{d} P_h(k)\nu^h \,. \tag{10}$$

That is, for $h, j = 0, 1, \ldots, d$, let $\lambda_{hj} \in \mathbb{F}_q$ be the Lagrange interpolation coefficients that satisfy

$$P_h(k) = \sum_{j=0}^{d} \lambda_{hj} P(\nu_j k)$$

for all $k \in K$. Observe in particular that the coefficients λ_{hj} depend only on $\nu_0, \nu_1, \ldots, \nu_d$, and can be computed once in $O(d^3 \mathrm{M}(q))$ time, for instance by using Gaussian elimination to solve the linear equation system consisting of (10) for all $\nu \in \{\nu_0, \nu_1, \ldots, \nu_d\}$.

With access to values of P_h at K, given a query $a \in \mathbb{F}_q^n$ we can use Lemma 2 to sieve for $P_h(a)$ for each $h = 0, 1, \ldots, d$. That is, we have

$$P(a) = \sum_{h=0}^{d} P_h(a) = -\sum_{h=0}^{d} \sum_{\tau \in \mathbb{F}_q^\times} \sum_{j=0}^{d} \tau^{q-1-rh} \lambda_{hj} P\big(\nu_j F(a,\tau)\big) \,.$$

The running time bounds follow from multiplying the running time bounds in Lemma 2 by d^2, as we use it that many times, after noting that the bound dominates the construction time of the λ_{hj} coefficients.

This completes the proof of Theorem 1. □

3.2 Proof of Theorem 2

Recall that s divides d and d/s divides $q-1$. Let $X_1, X_2, \ldots, X_d$ be the partition of variables for degree-separability. For $i = 1, 2, \ldots, s$, take

$$Y_i = X_{(i-1)d/s+1} \cup X_{(i-1)d/s+2} \cup \cdots \cup X_{id/s}$$

and observe that $|Y_i| = n/s$ for all i. Furthermore, observe that every monomial of a polynomial $P \in \mathbb{F}_q[x]$ that is degree-separable relative to $X_1, X_2, \ldots, X_d$ for every $i = 1, 2, \ldots, s$ has degree exactly d/s when restricted to the variables of Y_i.

Let us extend the construction in Lemma 1 into an s-fold product form over the partition $Y_1, Y_2, \ldots, Y_s$. Accordingly, we work with a multivariate polynomial over s indeterminates $\tau_1, \tau_2, \ldots, \tau_s$ instead of a univariate polynomial (3) over τ. Let $a = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{F}_q^n$ and let us write $a_{Y_i} \in \mathbb{F}_q^{n/s}$ for the restriction of a to coordinates in Y_i. Set $r = (q-1)s/d - 1$. Let us write $F_{Y_i}(a_{Y_i}, \tau_i) \in \mathbb{F}_q^n$ for the vector obtained by applying the construction given by (3) and (6) to the vector a_{Y_i} and τ_i, thereby obtaining a vector of length n/s indexed by Y_i, followed by padding with 0-entries outside the indices Y_i to obtain a vector of length n. Let us now define the (vector-valued) multivariate polynomial

$$F(a, \tau_1, \tau_2, \ldots, \tau_s) = F_{Y_1}(a_{Y_1}, \tau_1) + F_{Y_2}(a_{Y_2}, \tau_2) + \cdots + F_{Y_s}(a_{Y_s}, \tau_s)\,. \tag{11}$$

We observe by (3), (6), and (4) that $F(a, \tau_1, \tau_2, \ldots, \tau_s)$ ranges over all vectors of the form

$$\begin{aligned}\Bigg(&\left(\frac{\alpha_1}{r+1} + \tau_1\right)^{r+1} - \tau_1^{r+1}, \ldots, \left(\frac{\alpha_{n/s}}{r+1} + \tau_1\right)^{r+1} - \tau_1^{r+1},\\ &\left(\frac{\alpha_{n/s+1}}{r+1} + \tau_2\right)^{r+1} - \tau_2^{r+1}, \ldots, \left(\frac{\alpha_{2n/s}}{r+1} + \tau_2\right)^{r+1} - \tau_2^{r+1},\\ &\ldots,\\ &\left(\frac{\alpha_{n-n/s+1}}{r+1} + \tau_s\right)^{r+1} - \tau_s^{r+1}, \ldots, \left(\frac{\alpha_n}{r+1} + \tau_s\right)^{r+1} - \tau_s^{r+1}\Bigg)\end{aligned} \tag{12}$$

with $\alpha_1, \alpha_2, \ldots, \alpha_n, \tau_1, \tau_2, \ldots, \tau_s \in \mathbb{F}_q$. We define K to be the set of all such vectors. By similar reasoning as in the proof of Lemma 1, note that

$$|K| \le \left(\frac{q-1}{r+1} + 1\right)^{n+s} = \left(\frac{d}{s} + 1\right)^{n+s}\,.$$

Let $t = q - 1 - rd/s$ and observe that $t \ge 1$. From (7) and proceeding analogously as with the reasoning for (8) in the proof of Theorem 1, we thus have

$$P(a) = (-1)^s \sum_{\tau_1, \tau_2, \ldots, \tau_s \in \mathbb{F}_q^\times} \tau_1^t \tau_2^t \cdots \tau_s^t P(F(a, \tau_1, \tau_2, \ldots, \tau_s))\,. \tag{13}$$

Let $\beta_1, \beta_2, \ldots, \beta_{q-1}$ be an enumeration of the elements of $\mathbb{F}_q^\times$. For all $j = (j_1, j_2, \ldots, j_s) \in \{1, 2, \ldots, q-1\}^s$ take

$$g_j(a) = F(a, \beta_{j_1}, \beta_{j_2}, \ldots, \beta_{j_s}) \qquad \text{and} \qquad \gamma_j = (-1)^s \beta_{j_1}^t \beta_{j_2}^t \cdots \beta_{j_s}^t\,.$$

Theorem 2 now follows from (13). □

4 Fermionants

This section proves our two main theorems for evaluating fermionants. We start by noting that the fermionant is self-reducible, a result that easily follows from earlier work by Björklund [3], followed by the proofs of our present main theorems.

4.1 Self-Reducibility of the Fermionant

This subsection reviews how Björklund's [3] self-reducibility for permanents can be extended to fermionants. In essence, his methodology can be used to reduce the task of computing one fermionant of size $m \times m$ to the task of computing $2^{m-k}m^{O(1)}$ fermionants of size $k \times k$. We stress that this subsection is provided for ease of exposition only and no claim of originality is made.

The key idea is to view a fermionant as a sum over cycle covers, which enables the reduction from size $m \times m$ to size $k \times k$ by contracting cycles. That is, the actual reduction proceeds in the reverse direction from cycle covers of $[k]$ to cycle covers of $[m]$ by inserting walks and closed walks on $[m] \setminus [k]$ that are sieved with the principle of inclusion and exclusion so that only cycle covers remain.

Let us begin with basic terminology. For a set $S \subseteq [m]$ and an integer $\ell = 0, 1, 2, \ldots$, a *walk* of *length* ℓ on S is an $(\ell+1)$-tuple $(v_0, v_1, \ldots, v_\ell) \in S^{\ell+1}$. We say that the walk *starts* at v_0 and *ends* at v_ℓ. The walk is *closed* if $v_0 = v_\ell$ and $\ell \geq 1$; in this case we say that v_0 is the *root* of the walk. A walk is a *path* if $v_0, v_1, \ldots, v_\ell$ are distinct. A closed walk is a *cycle* if $v_0, v_1, \ldots, v_{\ell-1}$ are distinct and $v_0 < v_1, v_2, \ldots, v_{\ell-1}$ in the natural ordering of $[m]$. We say that a set of cycles $\{C_1, C_2, \ldots, C_c\}$ on S is a *cycle cover* of S if for all $u \in S$ it holds that u occurs in exactly one of the cycles $C_1, C_2, \ldots, C_c$. Let us write $\mathcal{C}^S$ for the set of all cycle covers of S.

Next we develop the fermionant as a sum over cycle covers. Let r, t, and a_{ij} for $i, j \in [m]$ be polynomial indeterminates and let $\mathbb{F}$ be the coefficient field. Associate with each walk $W = (v_0, v_1, \ldots, v_\ell) \in S^{\ell+1}$ the monomial

$$a(W) = \begin{cases} 1 & \text{if} \quad \ell = 0; \\ a_{v_0 v_1} a_{v_1 v_2} \cdots a_{v_{\ell-1} v_\ell} & \text{if} \quad \ell \geq 1. \end{cases} \tag{14}$$

The fermionant of $A = (a_{ij} : i, j \in [m])$ is now

$$\operatorname{fer}_t A = (-1)^m \sum_{\{C_1, C_2, \ldots, C_c\} \in \mathcal{C}^{[m]}} (-t)^c a(C_1) a(C_2) \cdots a(C_c)\,. \tag{15}$$

Let us now proceed with the reduction from size $m \times m$ to size $k \times k$ for $k = 0, 1, 2, \ldots, m$. Let us start with a high-level intuition. Consider an arbitrary cycle cover of $[m]$ with c cycles. The cycles in the cycle cover can be partitioned into two types:

1. cycles that contain at least one element of $[k]$, and
2. cycles that consist only of elements in $[m] \setminus [k]$.

We can now construct a cycle cover of $[k]$ as follows. First, transform each cycle W of Type 1 into a cycle W' on $[k]$ by deleting all the elements in $[m] \setminus [k]$ from W. Next, delete all cycles of Type 2. This results in a cycle cover of $[k]$. Observe that we can also proceed in the reverse direction; that is, we start from a cycle cover of $[k]$, insert vertex-disjoint paths on $[m] \setminus [k]$ in between the vertices on each cycle (that is, we transform W' to W in reverse), and finally insert vertex-disjoint cycles on $[m] \setminus [k]$ to complete a cycle cover of $[m]$. The reduction carries out this extension from a cycle cover of $[k]$ to a cycle cover of $[m]$ using walks and closed walks in $[m] \setminus [k]$. By applying the principle of inclusion and excusion over all $S \subseteq [m] \setminus [k]$, among the walks and closed walks all but paths and cycles will cancel in the sieve, and exactly cycle covers of $[m]$ remain.

Let us now proceed with the technical details of the reduction. First, we need a family of generating polynomials for walks. For $S \subseteq [m]\setminus[k]$, $i, j \in S$, and $\ell = 0, 1, \ldots, m$, consider the inductively-defined family of polynomials:

$$G^S_{\ell,i,j}(r) = \begin{cases} 1 & \text{if} \quad \ell = 0 \, and \, i = j; \\ 0 & \text{if} \quad \ell = 0 \, and \, i \neq j; \\ \sum_{u \in S} a_{iu} r G^S_{\ell-1,u,j}(r) & \text{if} \quad \ell \geq 1. \end{cases} \tag{16}$$

Let us write $\mathcal{W}^S_{\ell,i,j} \subseteq S^{\ell+1}$ for the set of all walks of length ℓ on S that start at i and end at j. The following lemma shows that the polynomials (16) indeed are a generating function for (edge-multisets of) walks.

Lemma 3 *We have the polynomial identity* $G^S_{\ell,i,j}(r) = r^\ell \sum_{W \in \mathcal{W}^S_{\ell,i,j}} a(W)$.

Proof Induction on $\ell = 0, 1, \ldots$ using (14) and (16). □

Second, let us develop a generating matrix for walks that will realize the Type-1 cycles. For $i, j \in [k]$ and $S \subseteq [m] \setminus [k]$, introduce the generating polynomial

$$\tilde{a}^S_{i,j}(r) = a_{ij} + \sum_{\ell=0}^{m-1} \sum_{u,v \in S} a_{iu} G^S_{\ell,u,v}(r) a_{vj} r \,. \tag{17}$$

Let us write $\hat{\mathcal{W}}^S_{i,j}$ for the set of walks that consists of (i) the walk (i, j) and (ii) all walks of the form $(i, v_0, v_1, \ldots, v_\ell, j)$ where $(v_0, v_1, \ldots, v_\ell)$ is a walk of length $0 \leq \ell \leq m-1$ on S. For a walk $W = (u_0, u_1, \ldots, u_\ell) \in [m]^{\ell+1}$ and a set $T \subseteq [m]$, let us write $|W \cap T| = |\{i \in \{0, 1, \ldots, \ell-1\} : u_i \in T\}|$ for the number of positions of W (not including the end position) that are in T.

Analogously to Lemma 3, the monomials of (17) track the edge-multisets of walks in $\hat{\mathcal{W}}_{i,j}$. Furthermore, the r-degree of each monomial records the number of positions on the walk in S.

Lemma 4 *We have the polynomial identity* $\tilde{a}^S_{i,j}(r) = \sum_{W \in \hat{\mathcal{W}}^S_{i,j}} r^{|W \cap S|} a(W)$.

Proof Use (17), Lemma 3, and the definition of $\hat{\mathcal{W}}^S_{i,j}$ to conclude identity. □

Let us arrange the coefficients $\tilde{a}^S_{i,j}(r)$ in (17) into a $k \times k$ matrix $\tilde{A}^S(r)$.

Third, let us develop a generating polynomial for closed walks that will realize Type-2 cycles. For $S \subseteq [m] \setminus [k]$ and $i \in S$, let us write

$$S_{\geq i} = \{u \in S : u \geq i\}\,.$$

For $S \subseteq [m] \setminus [k]$, $i \in S$, introduce the polynomial

$$H_i^S(r,t) = 1 - t \sum_{\ell=1}^{m} G_{\ell,i,i}^{S_{\geq i}}(r)\,. \tag{18}$$

The monomials of (18) track the edge-multisets of closed walks on $S_{\geq i}$ and rooted at i, including the possibility of no walk at all. The r-degree of each monomial records the length of the walk; by construction, the entire walk is on S. The t-degree determines whether a walk was made (degree 1) or not (degree 0). Also observe that the monomials that encode a walk occur with negative sign.

Fourth, introduce the polynomial

$$H^S(r,t) = \prod_{i \in S} H_i^S(r,t)\,. \tag{19}$$

Let us write $\mathcal{H}^S$ for the set of sets of closed walks on S such that (i) each closed walk has length ℓ with $1 \leq \ell \leq m$, (ii) the root of each closed walk is the minimum vertex on the closed walk, and (iii) no two walks in a set have the same root.

Lemma 5 *We have the polynomial identity*

$$H^S(r,t) = \sum_{\{D_1,D_2,\ldots,D_d\} \in \mathcal{H}^S} r^{\sum_{j=1}^{d} |D_j \cap S|}(-t)^d a(D_1)a(D_2)\cdots a(D_d)\,.$$

Proof Use (19), (18), Lemma 3, and the definition of $\mathcal{H}^S$ to yield identity. □

We are now ready to state and prove the main polynomial identity underlying the reduction. For a polynomial P in the indeterminate r, let us write $\{r^j\}P$ for the coefficient (polynomial) of the monomial r^j.

Theorem 4 *We have the polynomial identity*

$$\operatorname{fer}_t A = \left\{r^{m-k}\right\} \sum_{S \subseteq [m] \setminus [k]} (-1)^{|S|} H^S(r,t) \operatorname{fer}_t \tilde{A}^S(r)\,. \tag{20}$$

Proof Recall the cycle-cover form (15) of the fermionant. We show that the right-hand side of (20) reduces to the cycle-cover form. First, let us expand the inner $k \times k$ fermionant using (15) and (14) to obtain

$$\begin{aligned}\operatorname{fer}_t \tilde{A}^S(r) &= (-1)^k \sum_{\{\tilde{C}_1,\tilde{C}_2,\ldots,\tilde{C}_{\tilde{c}}\}\in\mathcal{C}^{[k]}} (-t)^{\tilde{c}} \prod_{j=1}^{\tilde{c}} \tilde{a}^S(\tilde{C}_j) \\ &= (-1)^k \sum_{\{\tilde{C}_1,\tilde{C}_2,\ldots,\tilde{C}_{\tilde{c}}\}\in\mathcal{C}^{[k]}} (-t)^{\tilde{c}} \prod_{j=1}^{\tilde{c}} \prod_{\substack{i=1\\ \tilde{C}_j=(u_0,u_1,\ldots,u_\ell)}}^{\ell} \tilde{a}^S_{u_{i-1}u_i}(r)\,.\end{aligned} \tag{21}$$

Let us write $\hat{\mathcal{C}}^{[k],S}$ for the set that consists of all sets $\{C_1, C_2, \ldots, C_{\tilde{c}}\}$ of closed walks in $[k] \cup S$ such that (i) when all the vertices in S are deleted from all the closed walks in a set, a cycle cover of $[k]$ results, and (ii) in each closed walk, there are at most m consecutive vertices in S. Now apply Lemma 4 to each $\tilde{a}^S_{u_{i-1}u_i}(r)$ in (21) to conlude that

$$\operatorname{fer}_t \tilde{A}^S(r) = (-1)^k \sum_{\{C_1,C_2,\ldots,C_{\tilde{c}}\}\in\hat{\mathcal{C}}^{[k],S}} (-t)^{\tilde{c}} \prod_{i=1}^{\tilde{c}} r^{|C_i\cap S|} a(C_i)\,. \tag{22}$$

Next multiply $H^S(r,t)$ with (22) and apply Lemma 4 to conlude that

$$\begin{aligned}&H^S(r,t)\operatorname{fer}_t \tilde{A}^S(r)\\ &= (-1)^k \sum_{\substack{\{C_1,C_2,\ldots,C_{\tilde{c}}\}\in\hat{\mathcal{C}}^{[k],S}\\ \{D_1,D_2,\ldots,D_d\}\in\mathcal{H}^S}} (-t)^{\tilde{c}+c} \prod_{i=1}^{\tilde{c}} r^{|C_i\cap S|} a(C_i) \prod_{j=1}^{d} r^{|D_j\cap S|} a(D_j)\,.\end{aligned} \tag{23}$$

Taking the coefficient of r^{m-k} in (23), we conclude that

$$\begin{aligned}&\left\{r^{m-k}\right\} H^S(r,t)\operatorname{fer}_t \tilde{A}^S(r)\\ &= (-1)^k \sum_{\substack{\{C_1,C_2,\ldots,C_{\tilde{c}}\}\in\hat{\mathcal{C}}^{[k],S}\\ \{D_1,D_2,\ldots,D_d\}\in\mathcal{H}^S\\ \sum_{i=1}^{\tilde{c}}|C_i\cap S|+\sum_{j=1}^{d}|D_j\cap S|=m-k}} (-t)^{\tilde{c}+d} \prod_{i=1}^{\tilde{c}} a(C_i) \prod_{j=1}^{d} a(D_j)\,.\end{aligned} \tag{24}$$

In particular, the sum in (24) is over exactly those $(\tilde{c}+d)$-sets of closed walks $\{C_1, C_2, \ldots, C_{\tilde{c}}, D_1, D_2, \ldots, D_d\}$ that (excluding their end-positions) in total have exactly $m-k$ positions that contain an element of S. Taking the $(-1)^{|S|}$-signed sum of (24) over all $S \subseteq [m] \setminus [k]$, by the principle of inclusion and exclusion we conclude that a set $\{C_1, C_2, \ldots, C_{\tilde{c}}, D_1, D_2, \ldots, D_d\}$ cancels unless it holds that for every element of $[m] \setminus [k]$ there is a unique position that contains the element. That is, together with (i) we have that $\{C_1, C_2, \ldots, C_{\tilde{c}}, D_1, D_2, \ldots, D_d\}$ forms a cycle cover of $[m]$. Moreover, each such cycle cover appears with sign $(-1)^{m-k}$. That is, from (24) and the principle of inclusion and exclusion, we have

$$\sum_{S\subseteq[m]\setminus[k]}(-1)^{|S|}\left\{r^{m-k}\right\}H^S(r,t)\,\mathrm{fer}_t\,\tilde{A}^S(r) =$$

$$= (-1)^m \sum_{\{C_1,C_2,\ldots,C_{\tilde{c}},D_1,D_2,\ldots,D_d\}\in\mathcal{C}^{[m]}} (-t)^{\tilde{c}+d}\prod_{i=1}^{\tilde{c}} a(C_i)\prod_{j=1}^{d} a(D_j)\,.$$

This agrees with the cycle-cover form (15) of the $m \times m$ fermionant $\mathrm{fer}_t\, A$ and thus establishes (20). □

Let us now develop the polynomial identity (20) into a scalar reduction from one $m \times m$ fermionant to a sum of $k \times k$ fermionants.

Theorem 5 *Suppose $|\mathbb{F}| \ge m^2 + 1$ and let $k = 0, 1, \ldots, m$. Then, there is an algorithm that given as input a matrix $A \in \mathbb{F}^{m\times m}$, a scalar $\tau \in \mathbb{F}$, and an integer $j = 1, 2, \ldots, 2^{m-k}(m^2 + 1)$, runs in time $m^{O(1)}$, executes $m^{O(1)}$ arithmetic operations in $\mathbb{F}$, and outputs a matrix $\tilde{A}_j \in \mathbb{F}^{k\times k}$ such that:*

$$\mathrm{fer}_\tau\, A = \sum_{j=1}^{2^{m-k}(m^2+1)} \mathrm{fer}_\tau\, \tilde{A}_j\,. \tag{25}$$

In particular, the fermionant $\mathrm{fer}_\tau\, A$ of a given $A \in \mathbb{F}^{m\times m}$ at $\tau \in \mathbb{F}$ can be computed in $2^m m^{O(1)}$ time and arithmetic operations in $\mathbb{F}$.

Proof Fix $k = 0, 1, \ldots, m$. Observe that the polynomial on the right-hand side of (20) has degree at most m^2 in the indeterminate r. Let $S_1, S_2, \ldots, S_{2^{m-k}}$ be an enumeration of all the 2^{m-k} subsets of $[m] \setminus [k]$ and let $\rho_0, \rho_1, \ldots, \rho_{m^2} \in \mathbb{F}$ be distinct.

Let $\lambda_0, \lambda_1, \ldots, \lambda_{m^2} \in \mathbb{F}$ be the unique Lagrange interpolation coefficients that for any polynomial $P(r) = \sum_{\ell=0}^{m^2}\pi_\ell r^\ell$ with coefficients $\pi_0, \pi_1, \ldots, \pi_{m^2} \in \mathbb{F}$ satisfy $\pi_{m-k} = \sum_{\ell=0}^{m^2}\lambda_\ell P(\rho_\ell)$. Observe that these coefficients can be computed in time and arithmetic operations in $\mathbb{F}$ bounded by a polynomial in m. From Theorem 4 it follows immediately that

$$\mathrm{fer}_\tau\, A = \sum_{i=0}^{2^{m-k}}\sum_{\ell=0}^{m^2}(-1)^{|S_i|}\lambda_\ell H^{S_i}(\rho_\ell,\tau)\,\mathrm{fer}_\tau\,\tilde{A}^{S_i}(\rho_\ell)\,. \tag{26}$$

Set up an efficiently computable bijection between $j = 1, 2, \ldots, 2^{m-k}(m^2 + 1)$ and the pairs of integers (i, ℓ) with $i = 1, 2, \ldots, 2^{m-k}$ and $\ell = 0, 1, \ldots, m^2$.

Given the matrix $A \in \mathbb{F}^{m\times m}$, the scalar $\tau \in \mathbb{F}$, and the integer $j = 1, 2, \ldots, 2^{m-k}(m^2 + 1)$ as input, the algorithm first computes the corresponding pair (i, ℓ) together with the subset S_i and the scalars λ_ℓ and ρ_ℓ. Next, with the substitutions $r \leftarrow \rho_\ell$ and $t \leftarrow \tau$, the algorithm uses dynamic programming on the recurrences (16), (18), (19), and (17) to compute the scalar $H^{S_i}(\rho_\ell, \tau) \in \mathbb{F}$ and the matrix $\tilde{A}^{S_i}(\rho_\ell) \in \mathbb{F}^{k\times k}$. This computation takes time and arithmetic operations in $\mathbb{F}$ bounded by a polynomial in m. Next the algorithm sets $A_j \in \mathbb{F}^{k\times k}$

equal to the matrix $\tilde{A}^{S_i}(\rho_\ell) \in \mathbb{F}^{k\times k}$ with all of the entries on its first row individually multiplied by the scalar $(-1)^{|S_i|}\lambda_\ell H^{S_i}(\rho_\ell, \tau)$. By (2), we have $\operatorname{fer}_\tau A_j = (-1)^{|S_i|}\lambda_\ell H^{S_i}(\rho_\ell, \tau) \operatorname{fer}_\tau \tilde{A}^{S_i}(\rho_\ell)$. The theorem now follows from (26) by taking the sum over $j = 1, 2, \ldots, 2^{m-k}(m^2+1)$. □

4.2 Proof of Theorem 3

Let $A \in \mathbb{F}_q^{m\times m}$ be given together with $\tau \in \mathbb{F}_q$. We seek to compute $\operatorname{fer}_\tau A$ and will deploy the self-reducibility enabled by Theorem 5 towards this end. By assumption we have that $q-1$ has a divisor u with $1.1\log q \le u \le 10\log q$. Since $m = \omega(\log^2 q \log\log q)$, for all large enough m we can let k be a multiple of u with

$$0.98\sqrt{m/\log\log q} \le k \le 0.99\sqrt{m/\log\log q}\,.$$

With the objective of applying Theorem 2, take $n = k^2$, $d = k$, and $s = k/u$. Observe that the fermionant (2) of a $k \times k$ matrix A at $\tau \in \mathbb{F}_q$ is a degree-separable polynomial P of degree d over the n variables in A. Furthermore, s divides d and d/s divides $q-1$, so the assumptions of Theorem 2 hold. By Theorem 5 we can evaluate this P at any given point (that is, for any given $k \times k$ matrix) in time $2^k k^{O(1)}$ and operations in $\mathbb{F}_q$. The tabulation of P for Theorem 2 thus can be done in time

$$\begin{aligned}
2^k k^{O(1)} \left(\frac{d}{s}+1\right)^{n+s} \mathrm{M}(q) &\le 2^k k^{O(1)} (u+1)^{0.99m/\log\log q+\sqrt{m}}\, \mathrm{M}(q)\\
&\le 2^k k^{O(1)} (20\log q)^{0.999m/\log\log q}\, \mathrm{M}(q)\\
&\le 2^{0.9999m}\mathrm{M}(q)\,.
\end{aligned}$$

Once the tabulation of P is complete, we can use the algorithms in Theorem 2 to query the $2^{m-k}m^{O(1)}$ fermionants of size $k \times k$ required by (25) in time $O(n(q-1)^s s\mathrm{M}(q))$ per query. Thus, the total time is at most

$$\begin{aligned}
&2^{m-k}q^s m^{O(1)}\mathrm{M}(q)\\
&\quad\le 2^{m-0.98\sqrt{m/\log\log q}}2^{(\log q)0.99\sqrt{m/\log\log q}/(1.1\log q)}m^{O(1)}\mathrm{M}(q)\\
&\quad= 2^{m-0.08\sqrt{m/\log\log q}}m^{O(1)}\mathrm{M}(q)\,.
\end{aligned}$$

This completes the proof of Theorem 3. □

4.3 Proof of Corollary 1

Here we show how to extend the algorithm to integers, via the Chinese Remainder Theorem. Let A be an integer matrix of size $m \times m$ with entries in $[-M, M]$ for $M = O(1)$. Let τ be an integer with $|\tau| = O(m)$. By Bertrand's postulate (e.g. [23, §I.1]) for all large enough m we can select a prime u with $5\log m \le u \le 10\log m$. Let us study the number of primes p in the interval $Mm^2 < p < Mm^4$ such that u divides

$p-1$. Let us write φ for Euler's totient function and recall the uniform variant of the Prime Number Theorem for arithmetic progressions [23, Corollary 8.31]. Namely, there is a constant $\gamma > 0$ such that, for any function $h(x)$ tending to infinity with x, and uniformly for $x \geq 3$ and $1 \leq u \leq (\ln x)^2/\bigl(h(x)^2(\ln\ln x)^6\bigr)$, we have

$$\sum_{\substack{p \leq x \\ p \equiv 1 \pmod{u}}} 1 = \frac{x}{\varphi(u)\ln x}\left(1 + O\left(\frac{1}{(\ln x)^{\gamma h(x)}}\right)\right). \tag{27}$$

Here the left-hand side sum in (27) is over all primes p at most x congruent to 1 modulo u.

Since u is prime, we have $\varphi(u) = u - 1 = \Theta(\log m)$. Thus from (27) we conclude that for all large enough m there exist at least $2m$ distinct primes p such that both $Mm^2 < p < Mm^4$ and u divides $p-1$. With the objective of satisfying the assumptions of Theorem 3, we conclude that u is in the interval $(1.1\log p, 10\log p)$ for these $2m$ primes p. Indeed, since M is a constant, for all large enough m we have $1.99\log m \leq \log p \leq 4.01\log m$, which implies $(5/4.01)\log p \leq 5\log m \leq u \leq 10\log m \leq (10/1.99)\log p$.

From (2) we observe that $|\operatorname{fer}_\tau A| \leq m! \cdot O(m)^m M^m < \frac{1}{2}m^{4m}M^{2m}$. Applying the Chinese Remainder Theorem together with Theorem 3 on A and τ over $\mathbb{F}_p$ for each of the $2m$ primes p in turn, we recover $\operatorname{fer}_\tau A$ over the integers, in time $2^{m-\Omega(\sqrt{m/\log\log m})}$. □

Acknowledgements This research was funded by the Swedish Research Council Grant VR 2016-03855 "Algebraic Graph Algorithms" (A.B.), the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement 338077 "Theory and Practice of Advanced Search and Enumeration" (P.K.), and the U.S. National Science Foundation under Grants CCF-1741638 and CCF-1741615 (R.W.). A preliminary conference abstract of this paper appeared in Daniel Lokshtanov and Naomi Nishimura (Eds.), Proceedings of the 12th International Symposium on Parameterized and Exact Computation (IPEC 2017, Vienna, Austria, September 4–8, 2017), LIPIcs: Leibniz International Proceedings in Informatics, Volume 89, Schloss Dagstuhl, Wadern, Germany, pp. 6.1–6.13, 2018.

References

1. Bax, E.T., Franklin, J.: A finite-difference sieve to count paths and cycles by length. Inf. Process. Lett. **60**(4), 171–176 (1996)
2. Besicovitch, A.S.: On Kakeya's problem and a similar one. Math. Z. **27**(1), 312–320 (1928)
3. Björklund, A.: Below all subsets for some permutational counting problems. In: Proceedings of the 15th SWAT, vol. 17, pp. 1–11 (2016)
4. Björklund, A., Husfeldt, T., Lyckberg, I.: Computing the permanent modulo a prime power. Inf. Process. Lett. **125**, 20–25 (2017)
5. Björklund, A., Kaski, P.: How proofs are prepared at Camelot: extended abstract. Proc. PODC **2016**, 391–400 (2016)
6. Björklund, A., Kaski, P., Koutis, I.: Directed Hamiltonicity and out-branchings via generalized Laplacians. In: Proceedings of the 44th ICALP, vol. 91, pp. 1–14 (2017)

7. Chandrasekharan, S., Wiese, U.: Partition functions of strongly correlated electron systems as "fermionants" (2011). arXiv:1108.2461v1
8. Cygan, M., Pilipczuk, M.: Faster exponential-time algorithms in graphs of bounded average degree. In: Proceedings of the 40th ICALP, pp. 364–375 (2013)
9. Dvir, Z.: From randomness extraction to rotating needles. Electron. Colloq. Comput. Complex. **16**(077), 1–19 (2009)
10. Dvir, Z.: Incidence theorems and their applications. (2012). arXiv:1208.5073
11. Gupta, A., Kamath, P., Kayal, N., Saptharishi, R.: Arithmetic circuits: a chasm at depth 3. SIAM J. Comput. **45**(3), 1064–1079 (2016)
12. Kedlaya, K.S., Umas, C.: Fast polynomial factorization and modular composition. SIAM J. Comput. **40**(6), 1767–1802 (2011)
13. Knuth, D.E.: The Art of Computer Programming, Volume 2: Seminumerical Algorithms. Addison-Wesley, Boston (1998)
14. Kopparty, S., Lev, V.F., Saraf, S., Sudan, M.: Kakyea-type sets in finite vector spaces. J. Algebr. Combin. **34**(3), 337–355 (2011)
15. Kyureghyan, G., Müller, P., Wang, Q.: On the size of Kakeya sets in finite vector spaces. Electron. J. Combin. **20**(3), 36 (2013)
16. Landsberg, J.M.: An introduction to geometric complexity theory. Eur. Math. Soc. Newsl. **99**, 10–18 (2016)
17. Lidl, R., Niederreiter, H.: Finite Fields, 2nd edn. Cambridge University Press, Cambridge (1997)
18. Lund, C., Fortnow, L., Karloff, H.J., Nisan, Noam: Algebraic methods for interactive proof systems. J. ACM **39**(4), 859–868 (1992)
19. Mertens, S., Moore, C.: The complexity of the fermionant, and immanants of constant width (2011). arXiv:1110.1821
20. Mockenhaupt, G., Tao, T.: Restriction and Kakeya phenomena for finite fields. Duke Math. J. **121**(1), 35–74 (2004)
21. Ryser, H.J.: Combinatorial Mathematics. Mathematical Association of America, Washington D.C. (1963)
22. Saraf, S., Sudan, M.: An improved lower bound on the size of Kakeya sets over finite fields. Anal. PDE **1**(3), 375–379 (2008)
23. Tenenbaum, G.: Introduction to Analytic and Probabilistic Number Theory, 3rd edn. American Mathematical Society, Washington D.C. (2015)
24. Valiant, L.G.: Completeness classes in algebra. In: Proceedings of the 11th STOC, pp. 249–261 (1979)
25. Valiant, L.G.: The complexity of computing the permanent. Theor. Comput. Sci. **8**, 189–201 (1979)
26. von zur Gathen, J., Gerhard, J.: Modern Computer Algebra, 3rd edn. Cambridge University Press, Cambridge (2013)
27. Williams, R.R: Strong ETH breaks with Merlin and Arthur: short non-interactive proofs of batch evaluation. In: Proceedings of the 31st CCC, vol. 2, pp. 1–17 (2016)
28. Wolff, T.: Recent work connected with the Kakeya problem. In: Rossi, H. (ed.) Prospects in Mathematics (Princeton, NJ, 1996), American Mathematical Society, Washington D.C., pp. 129–162 (1999)