Finite Matrix Multiplication Algorithms from Infinite Groups

Jonah Blasiak
Department of Mathematics, Drexel University, Philadelphia, PA, USA

Henry Cohn
Microsoft Research New England, One Memorial Drive, Cambridge, MA, USA

Joshua A. Grochow
Departments of Computer Science and Mathematics, University of Colorado Boulder, CO, USA

Kevin Pratt
Department of Computer Science, Courant Institute of Mathematical Sciences, New York, NY, USA

Chris Umans
Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA

Abstract

The Cohn–Umans (FOCS '03) group-theoretic framework for matrix multiplication produces fast matrix multiplication algorithms from three subsets of a finite group G satisfying a simple combinatorial condition (the Triple Product Property). The complexity of such an algorithm then depends on the representation theory of G. In this paper we extend the group-theoretic framework to the setting of *infinite* groups. In particular, this allows us to obtain constructions in Lie groups, with favorable parameters, that are provably impossible in finite groups of Lie type (Blasiak, Cohn, Grochow, Pratt, and Umans, ITCS '23). Previously the Lie group setting was investigated purely as an analogue of the finite group case; a key contribution in this paper is a fully developed framework for obtaining bona fide matrix multiplication algorithms directly from Lie group constructions.

As part of this framework, we introduce "separating functions" as a necessary new design component, and show that when the underlying group is $G = \mathrm{GL}_n$, these functions are polynomials with their degree being the key parameter. In particular, we show that a construction with "half-dimensional" subgroups and optimal degree would imply $\omega = 2$. We then build up machinery that reduces the problem of constructing optimal-degree separating polynomials to the problem of constructing a *single* polynomial (and a corresponding set of group elements) in a ring of invariant polynomials determined by two out of the three subgroups that satisfy the Triple Product Property. This machinery combines border rank with the Lie algebras associated with the Lie subgroups in a critical way.

We give several constructions illustrating the main components of the new framework, culminating in a construction in a special unitary group that achieves separating polynomials of optimal degree, meeting one of the key challenges. The subgroups in this construction have dimension *approaching* half the ambient dimension, but (just barely) too slowly. We argue that features of the classical Lie groups make it unlikely that constructions in these particular groups could produce nontrivial bounds on ω unless they prove $\omega = 2$. One way to get $\omega = 2$ via our new framework would be to lift our existing construction from the special unitary group to GL_n, and improve the dimension of the subgroups from $\frac{\dim G}{2} - \Theta(n)$ to $\frac{\dim G}{2} - o(n)$.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory

Keywords and phrases Fast matrix multiplication, representation theory, infinite groups

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.18

Related Version *Full Version*: `https://arXiv.org/abs/2410.14905` [10]

16th Innovations in Theoretical Computer Science Conference (ITCS 2025).
Editor: Raghu Meka; Article No. 18; pp. 18:1–18:23
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Funding *Jonah Blasiak*: Supported by NSF grant DMS-2154282.
Joshua A. Grochow: Supported by NSF CAREER award CCF-2047756.
Kevin Pratt: Supported by Subhash Khot's Simons Investigator Award.
Chris Umans: Supported by a Simons Foundation Investigator Award.

Acknowledgements We are grateful to Peter Bürgisser and Emma Church for useful discussions, and we thank the American Institute for Mathematics for hosting a SQuaRE, during which initial parts of this work were developed.

1 Introduction

Matrix multiplication is a fundamental algebraic operation with myriad algorithmic applications, and as such, determining its complexity is a central question in computational complexity. Since Strassen's 1969 discovery [26] that one could beat the straightforward $O(n^3)$ method with one that used only $O(n^{\log_2 7}) = O(n^{2.81\cdots})$ arithmetic operations, there has been a long line of work improving upper bounds on the complexity of matrix multiplication. It is standard to define the exponent ω of matrix multiplication as the smallest number such that $n \times n$ matrices can be multiplied using $n^{\omega+o(1)}$ arithmetic operations, and a somewhat surprising folklore conjecture is that $\omega = 2$, which would mean matrices can be multiplied asymptotically almost as quickly as they can be added. The current best bound is that $\omega \leq 2.371339$ [29], and proving or disproving that $\omega = 2$ remains a longstanding open question.

In [14], an approach towards this problem was proposed based on embedding matrix multiplication into the multiplication operation in the group algebra of a finite group. Group algebra multiplication then reduces to the multiplication of block-diagonal matrices, where the sizes of these blocks are determined by the (typically well-understood) representation theory of the group. Ultimately, one hopes to reduce a single matrix multiplication to the multiplication of (many) smaller matrices. Within this very general framework, certain families of groups have subsequently been shown to have structural properties that prevent a reduction that would yield $\omega = 2$ using this framework [7, 8, 9]. Other groups remain potentially viable, and the overall approach remains one of the two main lines of research towards improving upper bounds on ω, the other being the traditional "direct" tensor methods (e. g., [24, 27, 15, 28, 16, 21, 2, 18]), which also seem to be running up against several barrier results [3, 12, 1, 11].

In more detail, the reduction to group algebra multiplication is possible when one identifies a finite group G and sets $X, Y, Z \subseteq G$ satisfying the *triple product property* (TPP): for any $x, x' \in X$, $y, y' \in Y$, and $z, z' \in Z$,

$$xx'^{-1}yy'^{-1}zz'^{-1} = 1_G \iff x = x', y = y', z = z'.$$

If $X, Y, Z \subseteq G$ satisfy the TPP, then we can multiply two complex matrices A and B, of sizes $|X| \times |Y|$ and $|Y| \times |Z|$, resp., as follows. Index A with $X \times Y$, with $A[x, y]$ denoting the x, y entry of A, and index B with $Y \times Z$. Then we define the elements

$$\overline{A} = \sum_{x,y} A[x, y](xy^{-1}) \quad \text{and} \quad \overline{B} = \sum_{y,z} B[y, z](yz^{-1})$$

of the group ring $\mathbb{C}[G]$ and observe that the TPP implies that

$$\overline{A} \cdot \overline{B} = \sum_{x,z} (AB)[x, z](xz^{-1}) + E, \tag{1.1}$$

where $E \in \mathbb{C}[G]$ is supported on $XY^{-1}YZ^{-1} \setminus XZ^{-1}$. It is a standard fact of representation theory that, as algebras, $\mathbb{C}[G] \cong \bigoplus_i M_{d_i}(\mathbb{C})$, where M_{d_i} denotes the ring of $d_i \times d_i$ matrices, the sum is over the irreducible representations of G, and the d_i are the dimensions of those representations. This leads to the inequality:

▶ **Theorem 1.1** ([14, Theorem 4.1]). *If X, Y, Z satisfy the TPP in a finite group G, then*

$$(|X|\,|Y|\,|Z|)^{\omega/3} \leq \sum_i d_i^{\omega},$$

where d_i are the dimensions of the irreducible representations of G.

The upper bound on ω from Theorem 1.1 depends on a trade-off between the size of the matrix multiplication that can be embedded into $\mathbb{C}[G]$ (reflected in $|X|, |Y|, |Z|$) and the dimensions of the irreducible representations of G. In abelian groups, the latter are optimal: all d_i are 1. However, already in [14] it was observed that abelian groups cannot do better than the trivial construction $X = G$ and $Y = Z = \{1\}$, and thus cannot yield any bound better than the trivial bound $\omega \leq 3$. It was shown in [13] that non-abelian groups can achieve highly nontrivial bounds on ω (including many of the state-of-the-art bounds over the last decade), but obtaining such bounds requires a careful interplay between the size of the construction and the representation dimensions. Several families of groups have been shown not to admit constructions for which Theorem 1.1 would give $\omega = 2$, although many families of groups remain possibilities.

Because of the difficulty of finding such constructions, it is useful to look for other potential sources of examples of groups and group-like objects that could yield such constructions. Cohn and Umans [14] gave a construction of a TPP triple in the infinite group $\mathrm{GL}_n(\mathbb{R})$, despite having, at the time, no way of getting a (finite) matrix multiplication algorithm from such a construction.

One natural approach (that ends up not working) is to take a construction in a Lie group and try to transfer it to a finite group of Lie type, such as taking a construction in $\mathrm{GL}_n(\mathbb{R})$ and attempting an analogous construction in $\mathrm{GL}_n(\mathbb{F}_q)$. Indeed, a construction in $\mathrm{GL}_n(\mathbb{R})$ was given in [14] using the lower unitriangular, orthogonal, and upper unitriangular subgroups; this inspired a nontrivial TPP in the finite group $\mathrm{SL}_2(\mathbb{F}_q)$, where the orthogonal group is replaced by matrices of the form $\begin{pmatrix} 1+a & a \\ -a & 1-a \end{pmatrix}$. However, in that example, we have $|X| = |Y| = |Z| = q$, but $\mathrm{SL}_2(\mathbb{F}_q)$ has irreducible representations of dimension $q+1$, and reducing a $q \times q$ matrix multiplication to a $(q+1) \times (q+1)$ matrix multiplication doesn't give any bound on ω. More generally, in [9] it was shown that one cannot achieve $\omega = 2$ via Theorem 1.1 in any finite groups of Lie type, ruling out any such construction in $\mathrm{GL}_n(\mathbb{F}_q)$ or similar groups (though possibilities remain open to use related groups such as direct products of finite groups of Lie type). Thus, the Lie-type constructions remained only an analogy.

1.1 First main contribution: algorithms from infinite groups

In this paper, one of our main innovations is to extend the group-theoretic framework [14] to allow finite matrix multiplication algorithms from constructions in arbitrary – even infinite – groups. To achieve this, rather than using the entire group algebra $\mathbb{C}[G]$, which is infinite-dimensional when G is infinite, we focus only on sets of functions $G \to \mathbb{C}$ that (1) are linearly computable from a finite set of finite-dimensional representations of G (even when G is infinite) and (2) "separate" the elements in the group algebra that are in the linear span of $XY^{-1}YZ^{-1}$ (the support of (1.1)), in a sense made precise below (Definition 2.1). Our first main theorem in this generalized framework is then:

▶ **Theorem A** (Theorem 2.2). *Let G be a group (not necessarily finite), with finite subsets X, Y, Z satisfying the TPP. If R_{sep} is a set of finite-dimensional complex representations of G whose matrix entries separate $XY^{-1}YZ^{-1}$ (see Definition 2.1), then*

$$(|X|\,|Y|\,|Z|)^{\omega/3} \leq \sum_{\rho \in R_{\text{sep}}} (\dim \rho)^{\omega}.$$

It is even conceivable that this result could be used to improve the bounds from known constructions of TPPs in finite groups, by using only a subset of the group's irreducible representations rather than all of them.

But perhaps the main payoff of Theorem A is that it allows, for the first time, the derivation of matrix multiplication algorithms from infinite groups. This opens a huge variety of potential constructions to explore, even beyond those in Lie groups that will be the focus of the rest of the paper.

1.2 Second main contribution: quantitative targets for proving $\omega = 2$ in classical Lie groups

Another main contribution in this paper is to develop a series of tools and techniques, and to identify key targets to aim for, for getting good constructions using Theorem A in Lie groups such as $\mathrm{GL}_n(\mathbb{R})$, $\mathrm{GL}_n(\mathbb{C})$, or the unitary group U_n. Lie groups are defined as groups that are also smooth manifolds, and where the group multiplication and inverse are continuous in terms of the manifold topology. In fact, all of our constructions in this paper will be in one of these three groups, though much of the machinery we develop works for general matrix Lie groups, and it would be interesting to explore constructions in other Lie groups such as symplectic groups, orthogonal groups, exceptional simple Lie groups, or nilpotent or solvable Lie groups.

Before coming to the constructions, we highlight how the framework of Theorem A allows us to take the analogy from [14, 9], and turn it into a formal implication whose conclusion is a bound on ω. To briefly recall the analogy: elementary arguments show that any TPP triple in a finite group must satisfy $|X|\,|Y|\,|Z| \leq |G|^{3/2}$, called the *packing bound* because of the nature of the proof. Because $\sum d_i^2 = |G|$ in finite groups, it follows that any sequence of constructions that achieves $\omega = 2$ via Theorem 1.1 (not our new Theorem A) must asymptotically *meet the packing bound*, in the sense that $|X|\,|Y|\,|Z| \geq |G|^{3/2-o(1)}$. The analogy studied in [14, 9] is to think of Lie subgroups of dimension d as "roughly corresponding" to finite subsets of size q^d. Under this analogy, if X_n, Y_n, Z_n are families of Lie subgroups of Lie groups G_n that satisfy the TPP, then meeting the packing bound is analogous to the condition

$$\dim X_n + \dim Y_n + \dim Z_n \geq (3/2 - o_n(1)) \dim G_n. \tag{1.2}$$

A simple construction in the original Cohn–Umans paper [14, Theorem 6.1] shows this is indeed possible (with additional such constructions developed in [9]), and this construction forms the basis for a running example we will use throughout this paper to illustrate the development of our new framework.

Running example in $\mathbf{GL}_n(\mathbb{R})$

▶ **Theorem 1.2** ([14, Theorem 6.1]). *Let $G = \mathrm{GL}_n(\mathbb{R})$ and let X be the subgroup of lower unitriangular matrices, $Y = \mathrm{O}_n(\mathbb{R})$ (the orthogonal matrices), and Z be the subgroup of upper unitriangular matrices. Then the triple X, Y, Z satisfies the TPP.*

The dimension of G is n^2, and the dimension of each of the three subgroups is $n^2/2 - n/2$, so this construction meets the packing bound in the sense of (1.2) [9].

But there still remains the issue of how to use such a construction, even in concert with Theorem A, to get upper bounds on ω. And for this we require a deeper dive into representation theory, which will lead us to the key quantitative goals for these constructions.

In the case of the groups $\mathrm{GL}_n(\mathbb{R})$, $\mathrm{GL}_n(\mathbb{C})$, and U_n, rather than focusing on choosing arbitrary collections of representations to play the role of R_{sep} in Theorem A, we can exploit a relationship between the representations of these groups and the set of all degree-d polynomials. Namely, the irreducible representations of these groups are indexed by integer partitions into at most n parts, and the matrix entries of the representations corresponding to partitions of d, taken all together, span precisely the set of all degree-d polynomial functions on the group. Careful quantitative estimates then lead us to key targets for the degree of functions that separate out the elements of $XY^{-1}YZ^{-1}$ compared to the size of the finite TPP construction:

▶ **Theorem B** (Corollary 2.8, summarized). *If $X_{q,n}, Y_{q,n}, Z_{q,n} \subseteq \mathrm{GL}_n(\mathbb{C})$ (or U_n) satisfy the TPP and have sizes at least $q^{n^2/2-o_n(n)}$, and there are separating polynomials for $(X_{q,n}, Y_{q,n}, Z_{q,n})$ of degree at most $q^{1+o_q(1)}$, then Theorem A implies $\omega = 2$.*

Note how our new framework (Theorem A) has allowed us to take the analogy above whereby d-dimensional subgroups correspond to finite sets of size q^d, and turn it into a theorem that actually implies a bound on ω out of any such construction, rather than merely being an analogy. However, in this setup, as $\dim G = n^2$, we see that the bound we need on the size of the TPP triple is not $|X|\,|Y|\,|Z| \geq q^{(3/2-o_n(1))\dim G}$ as suggested in the previous Lie analogy of the packing bound [14, 9], but is slightly tighter, of the form $q^{(3/2-o_n(1/n))\dim G}$, and the example in Theorem 1.2 above falls short of this latter bound.

Our challenge is thus to find TPP constructions in GL_n or U_n that meet the bounds of Theorem B: three subsets in GL_n or U_n satisfying the TPP, of size at least $q^{n^2/2-o_n(n)}$, and admitting separating polynomials of degree at most $q^{1+o_q(1)}$.

1.3 Third main contribution: optimal degree using border rank, Lie algebras, and invariant theory

Our third main contribution is to show how to very nearly meet the conditions of Theorem B, using border rank, Lie algebras, and invariant theory. We also believe these techniques will have further uses. Our main theorem coming out of these constructions is:

▶ **Theorem C** (Summary of Theorem 3.1). *For any n and q, there are three subsets $X_q, Y_q, Z_q \subseteq \mathrm{U}_n$, all of size at least $q^{n^2/4-n/4}$, which satisfy the TPP and admit border-separating polynomials of degree $O(q)$.*

Note that this falls short of the conditions needed for Theorem B in only two ways: we get a construction of size $q^{(1/2)\dim G-\Theta(n)}$ where $G = \mathrm{U}_n$, whereas Theorem B would require both that the construction be in GL_n, and that it approach half the dimension just slightly faster, with sets of size $q^{(1/2)\dim \mathrm{GL}_n - o_n(n)}$, rather than our current $q^{(1/2)\dim \mathrm{U}_n - \Theta(n)}$. In addition to the sizes being very nearly right, the degree bound *does* satisfy the degree bound required by Theorem B (even slightly better than is needed: we get $O(q)$ whereas Theorem B only needs degree at most $q^{1+o_q(q)}$).

While in principle all one needs here is a sequence of finite sets X_q, Y_q, Z_q for infinitely many q, an appealing way to get such a sequence is to find Lie subgroups or submanifolds $X, Y, Z \subseteq \mathrm{GL}_n(\mathbb{C})$ – as in Theorem 1.2 – and then let X_q be some nicely constructed finite subset of X, etc. For example, when X is the subgroup of lower unitriangular matrices, we can take X_q to be the the lower unitriangular matrices with entries in $[0, q] \cap \mathbb{Z}$. And indeed, this is how the construction of Theorem C proceeds.

To get our constructions, we will combine three additional ingredients on top of Theorems A and B: Lie algebras, border rank, and invariant theory. Here we give a brief overview of these ingredients and how they mix together.

Lie algebras. In Sophus Lie's original development of Lie groups in the late 1800s, he realized that many questions about these continuous groups can be reduced to simpler questions of linear algebra, by focusing on the corresponding Lie algebras, which are, in particular, vector spaces, rather than more complicated manifolds. The Lie algebra $\mathrm{Lie}(G)$ associated to a Lie group G is "just" the tangent space to G (remember G is also a manifold) at the identity element. The Lie algebra is then a vector space. While the group structure of G induces an algebraic structure on $\mathrm{Lie}(G)$, in this paper we will have no need of its algebraic structure. We will only need a few basic facts (see, e. g., [22, 4, 19]):

- The Lie algebra of $\mathrm{GL}_n(\mathbb{C})$ is $M_n(\mathbb{C})$, the space of all $n \times n$ complex matrices.
- The Lie algebra of the orthogonal group $O_n(\mathbb{R})$ consists precisely of all skew-symmetric real matrices.
- The Lie algebra of the unitary group consists of the skew-Hermitian matrices.
- If G is a matrix Lie group – a Lie group that is a subgroup and submanifold of $\mathrm{GL}_n(\mathbb{C})$ – then its Lie algebra $\mathrm{Lie}(G)$ is a linear subspace of matrices. And if $A \in \mathrm{Lie}(G)$, then for all sufficiently small $\varepsilon > 0$, $\exp(\varepsilon A)$ (using the ordinary power series for the matrix exponential) is in G.

As with many problems on Lie groups, we would like to take advantage of the simpler, linear-algebraic nature of Lie algebras in our constructions.

Border rank. It turns out that the key tool for using Lie algebras in our setting is the concept of *border rank*. Although known to Terracini 100 years ago, border rank was rediscovered in the context of matrix multiplication by Bini *et al.* [5, 6]. Bini had developed computer code to search for algebraic algorithms for matrix multiplication, and some of the coefficients in his numerical calculations kept going off to infinity. At first he thought this was an error in his code, but in fact it reflects the fundamental phenomenon of border rank: for each fixed size n_0, it is possible that there is a sequence of algorithms, none of which correctly multiply $n_0 \times n_0$ matrices, but which, *in the limit*, in fact do so. If the algorithms in the sequence use only r non-constant multiplications, we say that matrix multiplication has border rank at most r. The border rank is always at most the ordinary rank, and it turns out that the exponent of matrix multiplication is the same whether measured with ordinary rank or border rank. Border rank has played an important role in essentially all newly developed matrix multiplication algorithm since then.

A bit more formally, a "border algorithm" for matrix multiplication can be viewed as a single bilinear algorithm that has coefficients that are Laurent series in ε – that is, power series that allow finitely many negative powers of ε as well – and such that it computes matrix multiplication in the limit as $\varepsilon \to 0$, that is, it computes a function of the form $(A, B) \mapsto AB + O(\varepsilon)$. (Note that, despite the algorithm itself being allowed to contain $1/\varepsilon$ in its intermediate operations, corresponding to Bini's coefficients that were going off to infinity, the function computed at the end should have such negative powers of ε cancel.)

Combining Lie algebras and border rank. It turns out that combining border rank with Lie algebras is very natural; here we exhibit just two advantages to doing so. If $X, Y, Z \subseteq \mathrm{GL}_n(\mathbb{C})$ are Lie subgroups that satisfy the TPP (as in Theorem 1.2), then we can take advantage of the simple linear-algebraic nature of their Lie algebras to help construct finite subsets of

X, Y, Z. An example we will return to is that if $Y = O_n(\mathbb{R})$ is the orthogonal group, it is a bit tricky to choose $q^{\dim O_n}$ many elements of Y in a principled way directly. But since $\mathrm{Lie}(Y)$ consists of all the skew-symmetric matrices, we can get a finite subset of Y by simply considering

$$Y'_\varepsilon = \{\exp(\varepsilon A) : A \text{ skew-symmetric with all } A_{ij} \in [0, q] \cap \mathbb{Z}\}$$

for $\varepsilon > 0$ sufficiently small.

A second advantage can be gotten by not choosing ε as above to be a fixed but small value, but rather allowing to the ε used in the expression $\exp(\varepsilon A)$ to be the same as the parameter ε used in the definition of border rank. In this setting, instead of finding a set of functions that exactly separates $XY^{-1}YZ^{-1}$ according to Definition 2.1, we can find functions that do so only up to $O(\varepsilon)$ (Definition 2.5). This both gives us more freedom in the construction, and combines very naturally with the Lie algebraic construction suggested above. Namely, when all the elements of X', Y', Z' are of the form $\exp(\varepsilon A)$ for various A in their Lie algebras, we see that an expression of the form $xy^{-1}y'z^{-1}$ becomes

$$\exp(\varepsilon A)\exp(-\varepsilon B)\exp(\varepsilon B')\exp(-\varepsilon C) = I + \varepsilon(A - B + B' - C) + O(\varepsilon^2).$$

Our border-separating polynomials can then directly access $A - B + B' - C$ by subtracting off I and dividing by ε; this leaves additional $O(\varepsilon)$ terms, but in the border setting that is still allowed. Thus combining Lie algebras and border rank lets us shift the problem from finding separating polynomials in the entries of a product of matrices and their inverses to finding (border-)separating polynomials in the entries of the simpler linear combination $A - B + B' - C$.

However, the linear combination $A - B + B' - C$ still "mixes" entries from the three subalgebras, and this causes some difficulty in the the task of designing (border-)separating polynomials. Our final ingredient is to use invariant theory to simplify this task even further.

1.4 Fourth main contribution: leveraging invariant polynomials

If we restrict our attention to polynomials $p(M)$ that are invariant under left multiplication by X and right multiplication by Z^{-1}, that is, $p(xMz^{-1}) = p(M)$ for all $x \in X, z \in Z$, we can get direct access to the $B' - B$ term above. Namely, if p is such an invariant polynomial, then

$$\begin{aligned} p(\exp(\varepsilon A)\exp(-\varepsilon B)\exp(\varepsilon B')\exp(-\varepsilon C)) &= p(\exp(-\varepsilon B)\exp(\varepsilon B')) \\ &= p(I + \varepsilon(-B + B') + O(\varepsilon^2)), \end{aligned}$$

where the first equality occurs because $\exp(\varepsilon A)$ is in X and $\exp(\varepsilon C)$ is in Z, and p is invariant under the action of X and Z. We codify this idea into the following key lemma, which is used in the proof of Theorem C. We state it in a simplified form, which is not correct as stated but captures the spirit of a special case. See Lemma 2.11 for the full and correct statement.

▶ **Lemma D** (Oversimplified version of Lemma 2.11). *Suppose X, Y, Z are three Lie subgroups of $\mathrm{GL}_n(\mathbb{C})$ that satisfy the TPP. If Y'_ε is a finite set of ε-parametrized families in Y, and there is a polynomial $p_\varepsilon(g)$ that is invariant under left multiplication by X and right multiplication by Z, such that*

$$p_\varepsilon(g) = \begin{cases} 1 + O(\varepsilon) & \text{if } g = I \\ 0 + O(\varepsilon) & \text{if } g \in (Y'_\varepsilon)^{-1}Y'_\varepsilon \setminus \{I\} \end{cases}$$

then there are finite subsets $X'_\varepsilon, Y'_\varepsilon$ of ε-parametrized elements of X and Z, of sizes $q^{\dim X}$ (resp., $q^{\dim Z}$) and border-separating polynomials for $(X'_\varepsilon, Y'_\varepsilon, Z'_\varepsilon)$ of degree $\deg p + O(q)$.

Thus, once we have a TPP construction of the "right" dimension (e. g., according to Theorem B), Lemma D implies that instead of finding appropriate finite subsets of X, Y, Z and a *set* of separating polynomials, to get $\omega = 2$ all we have to do is find an appropriate finite subset of Y and a *single* polynomial p as in the lemma, whose degree is $O(q)$.

When X and Z are well-studied subgroups, the ring of their invariant polynomials is often well understood, and this represents a significant simplification of the construction problem. The full detail of Lemma 2.11 is more complicated, but the additional complications in the statement of the lemma give us even further simplifications of the construction problem, which we take advantage of in our constructions in the running example and in Section 3. We also expect the techniques of this lemma to have further uses for additional constructions in the future.

1.5 Paper outline

In Section 2.1 we give the necessary definitions and proof of Theorem A, and in Section 2.2 we give its border rank version. In Section 2.3 we discuss the needed background on the representation theory of $\mathrm{GL}_n(\mathbb{C})$ – much of which we can use in a black-box fashion – and prove Theorem B. In Section 2.4 we show how to combine invariant theory, border rank, and Lie algebras to simplify the construction task; the key here is Lemma 2.11. In Section 3 we use the preceding ideas to give the construction that proves Theorem C. Finally, in Section 4 we conclude with our outlook and several open problems suggested by our framework.

Throughout the development of the framework in Section 2, we continue the example from Theorem 1.2 as a running example, and we show how each piece of the framework can be realized in that case.

The full version of the paper contains complete proofs of all claims [10].

2 A group-theoretic framework for infinite groups

In this section we describe the group-theoretic framework for obtaining matrix multiplication algorithms in infinite groups. The basic framework is in the next subsection, followed by a border-rank version for the important case of Lie groups. Subsection 2.3 proves some key bounds on the dimensions of irreducible representations of GL_n and U_n, which are the containing groups for our constructions in this paper. Finally subsection 2.4 describes machinery that reduces the main design task to finding a single separating polynomial in an invariant ring. These components come together in Section 3, where we give a construction achieving optimal degree separating polynomials.

2.1 Algorithms from the TPP in infinite groups

Let G be a group – not necessarily finite – with finite subsets X, Y, Z satisfying the TPP. Then we can embed matrix multiplication into the group algebra $\mathbb{C}[G]$ as in the case of finite groups. Note that $\mathbb{C}[G]$, where G may be infinite, consists of formal sums $\sum_{g \in G} \alpha_g g$, but now where each such sum has at most finitely many nonzero terms. Multiplication is as in the finite case.

To multiply a complex matrix A of size $|X| \times |Y|$ by a complex matrix B of size $|Y| \times |Z|$, we follow the approach described in the introduction. Indexing A by $X \times Y$ and B by $Y \times Z$, we define elements

$$\overline{A} = \sum_{x \in X, y \in Y} A[x, y](xy^{-1}) \quad \text{and} \quad \overline{B} = \sum_{y \in Y, z \in Z} B[y, z](yz^{-1})$$

of $\mathbb{C}[G]$ and observe that, as in the finite group case, the TPP implies that

$$\overline{A} \cdot \overline{B} = \sum_{x \in X, z \in Z} (AB)[x,z](xz^{-1}) + E, \tag{2.1}$$

where $E \in \mathbb{C}[G]$ is supported on $XY^{-1}YZ^{-1} \setminus XZ^{-1}$.

However, as the group algebra is no longer finite-dimensional when G is infinite, it is not immediately clear that the multiplication in the group algebra expressed in (2.1) can be carried out by a finite algorithm, even despite the fact that all the sums involved are themselves finite. One of our key new innovations is to introduce a new viewpoint on such a construction that will enable us to get a finite bilinear algorithm out of the above construction.

▶ **Definition 2.1.** *Given subsets $X, Y, Z \subseteq G$ of a group G, a* set of separating functions for (X, Y, Z) *is a collection of functions $\{f_{x,z} \colon G \to \mathbb{C} \mid x \in X, z \in Z\}$ such that*

$$f_{x,z}(g) = \begin{cases} 1 & \text{if } g = xz^{-1}, \text{ and} \\ 0 & \text{if } g \in XY^{-1}YZ^{-1} \setminus \{xz^{-1}\}. \end{cases}$$

Such functions always exist, but for them to be useful in matrix multiplication algorithms we need them to be in some sense "simple." To formulate the relevant notion – in which "simple" will ultimately imply "computable by smaller matrix multiplications" – we recall a standard definition from representation theory. Given a finite-dimensional representation $\rho \colon G \to \mathrm{GL}_n(\mathbb{C})$, a *representative function* of G associated to ρ (see, e. g., Procesi's book [23, §8.2][1]) is any function $G \to \mathbb{C}$ that is in the linear span of the functions $\{(g \mapsto \rho(g)_{i,j}) \mid i, j \in [n]\}$. Note that if ρ, ρ' are equivalent representations, they have the same set of representative functions. If $\mathcal{I}$ is a set of representations, we write $\mathrm{RepFun}(\mathcal{I})$ for the $\mathbb{C}$-linear span of all the representative functions of the representations in $\mathcal{I}$.

▶ **Theorem 2.2.** *Let G be a group (not necessarily finite), with finite subsets X, Y, Z satisfying the TPP. If R_{sep} is a finite set of finite-dimensional complex representations of G such that $\mathrm{RepFun}(R_{\mathrm{sep}})$ contains a set of separating functions for (X, Y, Z), then*

$$(|X|\,|Y|\,|Z|)^{\omega/3} \leq \sum_{\rho \in R_{\mathrm{sep}}} (\dim \rho)^{\omega}.$$

In particular, for $D = \sum_{\rho \in R_{\mathrm{sep}}} (\dim \rho)^2$ and $d_{max} = \max\{\dim \rho : \rho \in R_{\mathrm{sep}}\}$,

$$(|X|\,|Y|\,|Z|)^{\omega/3} \leq D \cdot d_{max}^{\omega-2}.$$

If G is a finite group, we may take $R_{\mathrm{sep}} = \mathrm{Irr}(G)$, the set of all irreducible representations of G, as $\mathrm{RepFun}(\mathrm{Irr}(G))$ is the collection of *all* functions $G \to \mathbb{C}$ when G is finite. In that case, we also have $D = |G|$, recovering the original theorem for finite groups [14, Theorem 4.1] as a special case of Theorem 2.2.

[1] The notion of representative function we use here is fairly standard; if one consults Procesi's book, one will see that his book works in the setting where G is a topological group, and requires the representations involved to be continuous, but the definition we use here works just as well. When G is a Lie group, as in our constructions later in the paper, the representations we use will in fact be continuous in the natural manifold topology on G rather than the discrete topology.

▶ Remark 2.3. In families of groups $\{G_i\}$ with "rapidly growing" irreducible representation dimensions, this theorem exhibits an "all-or-nothing" phenomenon, which we explain here. For the constructions in this paper and other natural constructions in classical Lie groups, $d_{max} \geq D^{1/2-g(i)}$ for some $g \in o(1)$ (and this is what we mean by "rapidly growing irreducible representation dimensions"). If we let $V = (|X||Y||Z|)^{1/3}$, it is clear that we must have $V > d_{max}$ for the inequality to imply *any* upper bound on ω. Together with the observation that d_{max} approaches $D^{1/2}$, this means that to prove any upper bound on ω we must have $V \geq D^{1/2-f(i)}$ for some $f \in o(1)$. The inequality from the theorem then becomes

$$D^{(1/2-f(i))\omega} \leq V^\omega \leq D \cdot d_{max}^{\omega-2} \leq D \cdot D^{(1/2-g(i))(\omega-2)}.$$

Taking the base-D logarithm of both sides, we get

$$\omega/2 - f(i)\omega \leq \omega/2 - g(i)\omega + 2g(i),$$

which implies $\omega \leq 2\left(\frac{g(i)}{g(i)-f(i)}\right)$. In constructions the natural thing to do is to ensure $f(i)$ goes to zero more rapidly than $g(i)$, in which case $\omega = 2$. While it is *possible* that $f(i)$ could approach $cg(i)$ for some $c > 0$ and yield an upper bound on ω strictly between 2 and 3, this would require detailed knowledge of the lower order terms of d_{max} and/or very fine control of the lower order terms of V, which we do not typically have.

Proof. Let X, Y, Z, R_{sep} be as in the statement, and let $\{f_{x,z} \mid x \in X, z \in Z\}$ be the claimed set of separating functions contained in $\text{RepFun}(R_{\text{sep}})$. For $f\colon G \to \mathbb{C}$, let $\overline{f}\colon \mathbb{C}[G] \to \mathbb{C}$ denote its unique linear extension to the group ring $\overline{f}(\sum \alpha_g g) := \sum \alpha_g f(g)$; as these sums have only finitely many nonzero terms by definition of the group ring (even when G is infinite), there is no issue of convergence. Applying $\overline{f_{x,z}}$ to (2.1) gives

$$\overline{f_{x,z}}(\overline{A} \cdot \overline{B}) = \sum_{x' \in X, z' \in Z} (AB)[x', z'] f_{x,z}(x'(z')^{-1}) + \overline{f_{x,z}}(E) = (AB)[x, z] + 0, \tag{2.2}$$

for $f_{x,z}$ is zero on all group elements of $X(Y)^{-1}Y(Z)^{-1}$ other than xz^{-1}, and (2.1) is entirely supported on $X(Y)^{-1}Y(Z)^{-1}$. To turn this into a finite algorithm, we will show that we can essentially do exactly the application of $\overline{f_{x,z}}$ in (2.2), but working only in the representations in R_{sep} rather than working in the full group ring.

For a representation $\rho \in R_{\text{sep}}$, let $\overline{\rho}$ denote the unique linear extension of ρ to $\mathbb{C}[G]$: $\overline{\rho}(\sum \alpha_g g) = \sum \alpha_g \rho(g)$, and let $\rho_{i,j}(g)$ be the (i,j) entry of the matrix $\rho(g)$, which we think of as a function $\rho_{i,j}\colon G \to \mathbb{C}$. Since $f_{x,z}$ is in $\text{RepFun}(R_{\text{sep}})$ by assumption, we can write $f_{x,z}$ as a $\mathbb{C}$-linear combination of the functions $\{\rho_{i,j} \mid \rho \in R_{\text{sep}},\ i,j \in [\dim \rho]\}$, say $f_{x,z} = \sum_{\rho \in R_{\text{sep}}} \sum_{i,j \in [\dim \rho]} M_{x,z,i,j} \rho_{i,j}$. Then we define $\widehat{f_{x,z}}(\rho)$ to be the matrix $M_{x,z,*,*}$, i.e.,

$$f_{x,z}(g) = \sum_{\rho \in R_{\text{sep}}} \sum_{i,j \in [\dim \rho]} \widehat{f_{x,z}}(\rho)_{i,j} \rho_{i,j}(g)$$

for all $x \in X, z \in Z, g \in G$. Finally, extending linearly and applying $\overline{f_{x,z}}$ to $\overline{A} \cdot \overline{B}$ as in (2.2), we get

$$\begin{aligned}
(AB)[x, z] &= \overline{f_{x,z}}(\overline{A} \cdot \overline{B}) \\
&= \sum_{\rho \in R_{\text{sep}}} \sum_{i,j \in [\dim \rho]} \widehat{f_{x,z}}(\rho)_{i,j} \overline{\rho_{i,j}}(\overline{A} \cdot \overline{B}) \\
&= \sum_{\rho \in R_{\text{sep}}} \langle \widehat{f_{x,z}}(\rho), \overline{\rho}(\overline{A} \cdot \overline{B}) \rangle \\
&= \sum_{\rho \in R_{\text{sep}}} \langle \widehat{f_{x,z}}(\rho), \overline{\rho}(\overline{A}) \cdot \overline{\rho}(\overline{B}) \rangle
\end{aligned}$$

The summation and inner product are linear functions whose coefficients are independent of the input matrices A, B, so they are "free" in a bilinear algorithm.

The product $\overline{\rho}(\overline{A}) \cdot \overline{\rho}(\overline{B})$ is a product of $d_\rho \times d_\rho$ matrices, where $d_\rho = \dim \rho$. Hence, the bilinear (i.e., tensor) rank of the preceding expression gives the following bound. Using rk to denote the tensor rank, and $\langle n, m, p\rangle$ to denote the tensor corresponding to matrix multiplication of an $n \times m$ matrix times and $m \times p$ matrix, we get

$$\operatorname{rk} \langle |X|, |Y|, |Z|\rangle \leq \sum_{\rho \in R_{\text{sep}}} \operatorname{rk} \langle d_\rho, d_\rho, d_\rho\rangle.$$

Exactly as in the finite group case [14], by symmetrizing we effectively get a square matrix multiplication of size $(|X|\,|Y|\,|Z|)^{1/3}$ on the left side, and by the tensor power trick the right side here can be replaced by $\sum_{\rho \in R_{\text{sep}}} (\dim \rho)^\omega$. For the last sentence of the theorem statement, we have $\sum_{\rho \in R_{\text{sep}}} d_\rho^\omega \leq \sum d_\rho^2 d_{max}^{\omega-2} = d_{max}^{\omega-2} \cdot D$. ◀

▶ Remark 2.4. The image $\overline{\rho}(\mathbb{C}[G])$ is the full $d_\rho \times d_\rho$ matrix ring if and only if ρ is an irreducible representation. In particular, although we will not take advantage of this in the present paper, we note that when ρ is not irreducible, we can replace $\operatorname{rk} \langle d_\rho, d_\rho, d_\rho\rangle$ (or d_ρ^ω) with the tensor rank of multiplying matrices in the image of $\overline{\rho}$, which may only be a subspace of all matrices. (This was true in the case of finite groups as well, but over $\mathbb{C}$ for finite groups we can use irreducible representations without loss of generality. For finite groups and representations in characteristic dividing $|G|$ this no longer holds, and for infinite groups even in characteristic zero it need not hold.)

2.2 Border rank version in Lie groups

For this section, let G be a (real or complex) Lie group. This includes familiar groups such as $\mathrm{GL}_n(\mathbb{R})$, $\mathrm{GL}_n(\mathbb{C})$, the orthogonal group O_n, and the unitary group U_n. Indeed, these examples will be the primary groups we use in our constructions later in the paper, although the framework laid out in this section is by no means limited to these particular examples.

For the purposes of this section, by a *1-parameter family* of elements of G we mean an analytic function $x\colon (0, \alpha) \to G$ for some $\alpha > 0$. If $X = \{x_1, \ldots, x_k\}$ is a collection of 1-parameter families x_i and ε is in their domain of definition, then we write $X(\varepsilon) = \{x_1(\varepsilon), \ldots, x_k(\varepsilon)\}$, which is just a finite subset of G.

▶ **Definition 2.5** (Border-separating functions). *Given sets X, Y, Z of 1-parameter families of elements of G with domain of definition $(0, \alpha)$, a* set of border-separating functions for (X, Y, Z) *is a collection of analytic functions $\{f_{x,z}\colon G \times (0, \alpha) \to \mathbb{C} \mid x \in X, z \in Z\}$ such that, for $0 < \varepsilon < \alpha$,*

$$f_{x,z}(g, \varepsilon) = \begin{cases} 1 + O(\varepsilon) & \text{if } g = x(\varepsilon)z(\varepsilon)^{-1}, \text{ and} \\ 0 + O(\varepsilon) & \text{if } g \in X(\varepsilon)Y(\varepsilon)^{-1}Y(\varepsilon)Z(\varepsilon)^{-1} \setminus \{x(\varepsilon)z(\varepsilon)^{-1}\}. \end{cases}$$

Here, as is standard for analytic functions of a small variable ε, the big-Oh notation means asymptotically as $\varepsilon \to 0$.

Now, we take our representative functions to come from *analytic* representations of our Lie group, and, as in most things related to border rank, allow ourselves 1-parameter families of representative functions. Let $\mathbb{C}^{(0,\alpha)}$ denote the set of analytic functions $(0, \alpha) \to \mathbb{C}$. Given a set $\mathcal{I}$ of analytic representations of G, we write $\mathrm{RepFun}_{\text{fam}}(\mathcal{I})$ for the $\mathbb{C}^{(0,\alpha)}$-linear span of all the representative functions associated to any $\rho \in \mathcal{I}$.

Given three sets X, Y, Z of 1-parameter families of elements of G, we say they satisfy the TPP if $X(\varepsilon), Y(\varepsilon), Z(\varepsilon)$ satisfy the TPP for all ε in their domain of definition. Given such X, Y, Z, we may encode ε-approximate matrix multiplication (à la border rank) into the group ring $\mathbb{C}[G]$ in a similar manner as before, but now parameterized by $\varepsilon \in (0, \alpha)$. For an $|X| \times |Y|$ matrix A and $|Y| \times |Z|$ matrix B, we define the following functions $(0, \alpha) \to \mathbb{C}[G]$:

$$\overline{A}(\varepsilon) = \sum_{i \in [n], j \in [m]} A[i,j](x_i(\varepsilon) y_j(\varepsilon)^{-1}) \quad \text{and} \quad \overline{B}(\varepsilon) = \sum_{j \in [m], k \in [\ell]} B[j,k](y_j(\varepsilon) z_k(\varepsilon)^{-1}).$$

Then

$$\overline{A}(\varepsilon) \cdot \overline{B}(\varepsilon) = \sum_{i \in [n], k \in [\ell]} (AB)[i,k](x(\varepsilon) z(\varepsilon)^{-1}) + E(\varepsilon), \tag{2.3}$$

where $E(\varepsilon) \in \mathbb{C}[G]$ is supported on $X(\varepsilon)Y(\varepsilon)^{-1}Y(\varepsilon)Z(\varepsilon)^{-1} \setminus X(\varepsilon)Z(\varepsilon)^{-1}$.

▶ **Theorem 2.6.** *Let G be a* ***Lie***[2] *group, with finite sets X, Y, Z of* ***1-parameter families*** *satisfying the TPP. If R_{sep} is a finite set of finite-dimensional* ***analytic*** *representations of G such that $\text{RepFun}_{\boldsymbol{fam}}(R_{\text{sep}})$ contains a set of* ***border-****separating functions for (X, Y, Z), then*

$$(|X|\,|Y|\,|Z|)^{\omega/3} \leq \sum_{\rho \in R_{\text{sep}}} (\dim \rho)^{\omega}.$$

In particular, for $D = \sum_{\rho \in R_{\text{sep}}} (\dim \rho)^2$, and $d_{max} = \max\{\dim \rho : \rho \in R_{\text{sep}}\}$,

$$(|X|\,|Y|\,|Z|)^{\omega/3} \leq D \cdot d_{max}^{\omega-2}.$$

The proof is very similar to Theorem 2.2, but using border rank instead of rank, so we postpone it to the full version of the paper [10].

An appealing way to use the freedom of border rank is to first find a TPP construction of Lie subgroups X, Y, Z, and then to define finite subsets of those three using their Lie algebras, viz.:

$$X' = \{\exp(\varepsilon a) : a \in A, \text{ some finite subset of the Lie algebra of } X\},$$

where $\varepsilon \to 0$ is the parameter we use for border rank. We will in fact show how to do this in a fairly generic way in Lemma 2.11 below, using some additional machinery that we develop first.

2.3 From representations to degree bounds in GL_n and U_n

In this paper, our constructions will all take place in GL_n or (slight variants of) the unitary group U_n, although the framework of Theorems 2.2 and 2.6 is much more general. These groups have some useful properties that will motivate some of the machinery we develop below – even though that machinery will also end up being more general – so we take a brief interlude to highlight the relevant properties of GL_n, before returning to the general abstract framework in the next section.

For both GL_n and U_n, there is a natural correspondence between representations and polynomials of a given degree. By focusing on separating polynomials (instead of more general separating functions), this correspondence will allow our constructions to focus only

[2] We have put in **bold** the parts of the statement of Theorem 2.6 that differ from Theorem 2.2.

on the *degree* of the separating polynomials. The upshot is that instead of thinking directly about what representations will comprise R_{sep}, we can focus solely on the degree of our separating polynomials when working in these groups.

In the rest of this section we review the relevant aspects of the representation theory of GL_n and U_n, and extract from those the relevant target bounds on degree to get bounds on ω. A standard reference for the representation theory of these groups is [19]. It is a standard fact in representation theory that the finite-dimensional representation theory of these two groups are essentially the same, and everything we say in this section will apply to both groups.

The irreducible polynomial representations ρ of $\mathrm{GL}_n(\mathbb{C})$ and U_n – where the entries of the matrix $\rho(g)$ are polynomials in the entries of the matrix g – are indexed by integer partitions $\lambda = (\lambda_1, \ldots, \lambda_n)$ with at most n parts, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. The matrix coefficients of the irreducible representations indexed by partitions of s span the functions from G to $\mathbb{C}$ that are expressible as degree s polynomials in the entries of GL_n. We write $\mathrm{Irr}_s(G)$ to denote the set of (pairwise non-isomorphic) irreducible representations of G indexed by partitions of the integer s.

The following bound will determine the target degree of the separating polynomials in our constructions, as in Corollary 2.8 below.

▶ **Lemma 2.7.** *Let $n \geq 3$ and $s \geq 2$. For $\rho \in \mathrm{Irr}_s(\mathrm{GL}_n)$ (resp., $\mathrm{Irr}_s(\mathrm{U}_n)$), $\dim(\rho) \leq s^{\binom{n}{2}}$.*

Although the result follows from some simple asymptotic analysis of standard results about the representation theory of GL_n, we could not find it in the literature, so we provide a full proof in [10].

Here and below it is helpful to think of n as large but fixed, and s growing independently of n.

▶ **Corollary 2.8.** *Suppose $X, Y, Z \subseteq \mathrm{GL}_n(\mathbb{C})$ (or U_n) satisfy the TPP, and there is a set of separating polynomials for (X, Y, Z) of degree at most s. Then ω satisfies*

$$(|X|\,|Y|\,|Z|)^{\omega/3} \leq s^{\binom{n}{2}(\omega-2)} \cdot \binom{s+n^2}{n^2}.$$

In particular, if $|X|, |Y|, |Z| \geq s^{(1-o_s(1))(n^2/2-o_n(n))}$, then $\omega = 2$.

The same holds if X, Y, Z are 1-parameter families that satisfy the TPP and we replace separating polynomials with border-separating polynomials.

Equivalently, if $|X|, |Y|, |Z| \geq q^{n^2/2-o_n(n)}$ and there are (border-)separating polynomials of degree at most $q^{1+o_q(1)}$, then $\omega = 2$. See the full version of the paper for a proof [10].

Next we show that the Lie TPP construction of [14] (Theorem 1.2 above) in fact admits separating polynomials of degree $O(q^2)$. In Section 2.4, we will show how to use a border-rank version of our framework to construct *finite* sets X, Y, Z, whose size we can then compare to the degree of the separating polynomials as in Corollary 2.8.

Separating polynomials for the running example in $\mathbf{GL}_n(\mathbb{R})$

We begin by describing a finite subset Y_q of the orthogonal group O_n, with the property that the diagonal entries of matrices in the quotient set $Q(Y_q)$ have a limited number of possible values, in the sense formalized in the following lemma. Here we treat n as fixed, so the implicit constant in $O(q^2)$ can depend on n.

▶ **Lemma 2.9.** *For each positive integer q, there exists a subset $Y_q \subseteq \mathrm{O}_n$ of cardinality at least $q^{n^2/2-(5/2)n}$, with the following property: for all $y, y' \in Y_q$, if y and y' agree in their first i columns, then*

$$(y^T y')[i+1, i+1] \in W_q,$$

where $W_q \subseteq \mathbb{R}$ is a fixed finite set of cardinality $O(q^2)$.

For a proof, see the full version of the paper [10].

We remark that one might hope to improve the cardinality of W_q to $O(q^{2-\varepsilon})$ by replacing V_i with a more cleverly chosen set of unit vectors with fewer distinct pairwise inner products. This would lead to an improved bound on the degree of the separating polynomials described in the following theorem, and if one could take $\varepsilon = 1$ this would yield separating polynomials of optimal degree. Unfortunately, this is impossible by the solution to the high-dimensional version of the Erdős distinct distances problem [25].

▶ **Theorem 2.10.** *Let G, X, Y, Z be as in Theorem 1.2, let $X_q \subseteq X$ be those lower unitriangular matrices whose below-diagonal entries are integers in $\{1, 2, \ldots, q\}$, let $Z_q \subseteq Z$ be those upper unitriangular matrices whose above-diagonal entries are integers in $\{1, 2, \ldots, q\}$, and let $Y_q \subseteq Y$ be the set of orthogonal matrices guaranteed by Lemma 2.9. Then there are separating polynomials for X_q, Y_q, Z_q of degree $O(q^2)$.*

For a proof, see the full version of the paper [10].

Note that here we have $\dim X = \dim Y = \dim Z = (n^2 - n)/2$, but for Corollary 2.8 we would need them to have dimension $n^2/2 - o_n(n)$, and the separating polynomial we get has degree $O_q(q^2)$, but Corollary 2.8 needs degree at most $q^{1+o_q(1)}$. In Section 3 we give a construction of the same relative dimension (namely, $\frac{\dim G}{2} - \Theta(n)$) but with separating polynomials of degree $O(q)$, meeting the degree bound of Corollary 2.8.

2.4 Separating polynomials from a single invariant polynomial

In this section, we return to the case of a general matrix Lie group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ or $\mathrm{GL}_n(\mathbb{R})$, and we show how to leverage invariant theory to construct separating polynomials whose degree is not too large, and that thus have a hope of meeting the degree bound of Corollary 2.8 in the case of $G = \mathrm{GL}_n(\mathbb{C})$ or $G = \mathrm{U}_n$.

The key result in this section, Lemma 2.11, combines border rank and Lie algebras, as suggested at the end of Section 2.2, with a new ingredient from invariant theory. The lemma lets us "split" the construction of sets satisfying the TPP and (border-)separating polynomials into two parts: the first part only involves the first and third sets X and Z (essentially, a separating polynomial version of the "double product property" (DPP), which is that $x^{-1}x'z^{-1}z' = 1$ iff $x = x'$ and $z = z'$)), and the second part involves only the middle set Y and the ring of XZ-invariant polynomials. From one view, this lets us start with a Y, and "work backwards," shifting the design task onto X, Z, and their invariant polynomials.

More specifically, Lemma 2.11 instantiates the following idea. Roughly, we would like to construct the separating polynomial $f_{x,z}$ as a product of three factors: one that is the indicator function of x among the set X (1 on x, zero on all other elements of X), one that is the indicator function of z^{-1} among Z^{-1}, and one on that is the indicator function p_0 of I among $Y^{-1}Y$. However, the issue here is that $f_{x,z}$, and hence these factors, does not really have separate access to these three parts: it only gets as input the product $xy^{-1}y'z^{-1}$. Our initial motivation to use invariant polynomials was that if p_0 has the property that $p_0(xy^{-1}y'z^{-1}) = p_0(y^{-1}y')$ for all $x \in X, z \in Z$, then it in fact *does* get direct access to only the Y-part of the input. Lemma 2.11 then gives generic machinery that, from such an XZ-invariant p_0, constructs a full set of (border-)separating polynomials, which have the correct degree if p_0 has the correct degree.

One advantage of this approach, in addition to separating out the design task associated with Y more from the specification of X and Z, is that it only necessitates the construction of at most three polynomials p_0, p_X, p_Z, rather than a whole set $\{f_{x,z}\}$ of $|X|\,|Z|$-many polynomials, and the p_X, p_Z polynomials are usually easy to construct if X and Z satisfy the DPP.

We now proceed with the technical details. By a *polynomial* on a matrix Lie group $G \subseteq \mathrm{GL}_n(\mathbb{C})$ or $G \subseteq \mathrm{GL}_n(\mathbb{R})$, we mean a function $G \to \mathbb{C}$ that can be expressed as a polynomial in the n^2 matrix entries x_{ij} of elements of G, with complex coefficients,[3] and similarly for a polynomial on the Lie algebra $\mathrm{Lie}(G) \subseteq M_n(\mathbb{C})$. By a 1-parameter family of polynomials we mean a polynomial in the n^2 matrix entries x_{ij} of elements of G, with coefficients in $\mathbb{C}^{(0,\alpha)}$ for some $\alpha > 0$.

Because we will use it frequently, we introduce the notation

$$\mathrm{Inv}_{X,Z} = \{p \in \mathbb{C}[M_{ij} : i,j \in [n]] \mid p(xMz) = p(M) \text{ for all } x \in X, z \in Z\}.$$

It is readily observed that $\mathrm{Inv}_{X,Z}$ is closed under multiplication and addition, and hence is a subring of the polynomial ring. In favorable (and many well-studied) cases, this subring is finitely generated over $\mathbb{C}$ (see, e. g., [20, 17]), and when a finite generating set of $\mathrm{Inv}_{X,Z}$ is known, we can focus on designing a polynomial that is composed with those generating invariants – which is then automatically invariant itself – in order to obtain p_0. We will see an example of this below.

In a Lie group G, for convenience let us say a subset $X \subseteq G$ is a *Lie submanifold* if X is a submanifold containing the identity of G, and such that for every v in the tangent space of X at the identity, for all sufficiently small $\varepsilon > 0$, the exponential $\exp(\varepsilon t)$ is in X. If X is a Lie submanifold, we use $\mathrm{Lie}(X)$ to denote the tangent space to X at the identity, even though this need not be a Lie algebra.

▶ **Lemma 2.11** (Splitting separating functions into invariant functions and double-product property)**.** *Let G be a matrix Lie subgroup of $\mathrm{GL}_n(\mathbb{F})$ for $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, with Lie submanifolds X, Z and a finite set Y of 1-parameter families. Let q be a positive integer.*

[3] Warning: for readers familiar with algebraic geometry, if G is an algebraic group our definition here is *not* quite the same thing as an element of the coordinate ring of G. For example, when $G = \mathrm{U}_n$, we view G as a subset of $\mathrm{GL}_n(\mathbb{C})$ and want to consider polynomials only in the n^2 matrix entries of these complex matrices, but as an algebraic group G is a *real* algebraic group, with twice as many (real) coordinates, and this difference of a factor of 2 can actually be important. We believe there is a modification of our framework that works in the setting of algebraic groups, but leave that for future work.

If there exist

- *continuous functions* $f_X\colon \mathbb{F}^{d_X} \to \mathrm{Lie}(X)$ *and* $f_Z\colon \mathbb{F}^{d_Z} \to \mathrm{Lie}(Z)$*, and polynomials* $p_X\colon \mathrm{Lie}(G) \to \mathbb{F}^{d_X}$ *and* $p_Z\colon \mathrm{Lie}(G) \to \mathbb{F}^{d_Z}$ *such that*

$$p_X(f_X(v) - f_Z(v')) = v \qquad \text{and} \qquad p_Z(f_X(v) - f_Z(v')) = v',$$

for all $v \in \mathbb{F}^{d_X}, v' \in \mathbb{F}^{d_Z}$,[4]

and

- *a 1-parameter family of polynomials* $p_0(g,\varepsilon) \in \mathrm{Inv}_{X,Z}$ *such that*

$$p_0(g,\varepsilon) = \begin{cases} 1 + O(\varepsilon) & \text{if } g = I \\ 0 + O(\varepsilon) & \text{if } g \in Y(\varepsilon)^{-1}Y(\varepsilon) \setminus \{I\}, \end{cases}$$

then there exist

- *a reparametrization* Y' *of* Y *and finite sets* X', Z' *of 1-parameter families in* X*,* Z *(resp.) such that* $X'(\varepsilon), Y'(\varepsilon), Z'(\varepsilon)$ *satisfy the TPP for all* ε *in some non-empty range* $(0,\alpha')$*,* $|X'| \geq q^{d_X}$*, and* $|Z'| \geq q^{d_Z}$*, and*
- *a set of border-separating polynomials for* (X', Y', Z') *of degree*[5] *at most* $\deg p_0 + O((d_X + d_Z)q)$*.*

As the first condition of Lemma 2.11 may be a little intimidating, before the proof we give an example to help allay such fears:

▶ **Lemma 2.12.** *If* $X, Z \leq G \leq \mathrm{GL}_n(\mathbb{R})$ *are Lie subgroups that intersect trivially, then there exist* f_X, f_Z, p_X, p_Z *satisfying the first condition of Lemma 2.11, with* $d_X = \dim X$ *and* $d_Z = \dim Z$*.*

Note that if we drop the requirement that p_0 be in $\mathrm{Inv}_{X,Z}$, then the hypotheses of Lemma 2.11 would be almost trivially satisfied whenever X and Z have trivial intersection. In applying Lemma 2.11, it is only the use of invariants that adds any real difficulty; however, the use of invariants is also a lynchpin to reach the conclusion.

For the proofs of Lemmas 2.11 and 2.12. see the full version of the paper [10].

Continuing our running example, we now show how to satisfy the hypothesis of Lemma 2.11, and thus get a TPP construction using border rank, invariant polynomials, and Lie algebras.

Example of a separating polynomial in the invariant ring

We return to the example in $\mathrm{GL}_n(\mathbb{R})$ described in Theorem 1.2. The following lemma is straightforward to check. Let lpm_i denote the polynomial that is the i-th leading principal minor of an $n \times n$ matrix; that is, lpm_i is the determinant of the upper-left $i \times i$ submatrix of an $n \times n$ matrix.

▶ **Lemma 2.13.** *Let* $X, Z \subseteq \mathrm{GL}_n(\mathbb{R})$ *be the sets of lower unitriangular matrices and the upper unitriangular matrices, respectively. Then* $\mathrm{Inv}_{X,Z}$ *contains the leading principal minors.*

---

[4] Notice that, since p_X, p_Z essentially invert f_X, f_Z on the sum of their images, this condition implies that $\mathrm{Span}_{\mathbb{F}}(\mathrm{Image}(f_X)) \cap \mathrm{Span}_{\mathbb{F}}(\mathrm{Image}(f_Z)) = 0$, a DPP-like condition.

[5] Degree as measured only as a function of the matrix entries – ε is still considered a separate parameter. Equivalently, degree as polynomials with coefficients in $\mathbb{C}^{(0,\alpha)}$.

In fact, in this case $\mathrm{Inv}_{X,Z}$ is generated by the leading principal minors, but we will not need this stronger fact for our constructions.

▶ **Theorem 2.14.** *Let G, X, Y, Z be as in Theorem 1.2. For any $q > 0$, there is a set $\widetilde{Y}$ of 1-parameter families taking values in Y, of size $q^{\dim Y}$, as well as functions f_X, f_Z, polynomials p_X, p_Z, and a 1-parameter family of polynomials p_0 of degree $O(q^2)$ satisfying the hypotheses of Lemma 2.11 with $d_X = \dim X$, $d_Z = \dim Z$, and $\widetilde{Y}$ replacing Y.*

Consequently, there is a TPP construction with sizes $q^{\dim X}, q^{\dim Y}, q^{\dim Z}$, resp., admitting border-separating polynomials of degree $O(q^2)$.

For a proof, see the full version of the paper [10].

3 Separating polynomials of degree $O(q)$

In this section we give a construction, now in a non-compact special unitary group, in which the Lie subgroups X, Y, Z satisfy the TPP and have dimension approaching half the ambient dimension. We show the existence of finite sets $X_q \subseteq X$, $Y_q \subseteq Y$, and $Z_q \subseteq Z$ satisfying $|X_q| = q^{\dim X}$, $|Y_q| = q^{\dim Y}$, $|Z_q| = q^{\dim Z}$ and with separating polynomials of degree $O(q)$. As shown in Corollary 2.8, this degree bound is the quantitative requirement on the degree for achieving exponent 2; where the construction falls short is that it approaches half the dimension "just barely" too slowly, at a rate of $\Theta_n(n)$ rather than $o_n(n)$, and second, it is in a unitary group, so it approaches half the dimension of the unitary group, rather than half the dimension of $\mathrm{GL}_n(\mathbb{C})$.

3.1 The construction

We assume n is even and define the containing group

$$G = \mathrm{SU}_{n/2,n/2} = \{M \in \mathrm{GL}_n(\mathbb{C}) \mid M^*QM = Q \text{ and } \det(M) = 1\},$$
$$\text{with } Q = \left(\begin{array}{c|c} I & \\ \hline & -I \end{array}\right).$$

Further, let J be the matrix with ones on the antidiagonal (in positions $(n+1-i, i)$),

$$U = \frac{1}{\sqrt{2}} \cdot \left(\begin{array}{c|c} I & J \\ \hline J & -I \end{array}\right),$$
$$D_0 = \mathrm{diag}(n, n-1, n-2, \ldots, n/2+1, (n/2+1)^{-1}, (n/2+2)^{-1}, \ldots, n^{-1}),$$

and

$$D = UD_0U^*. \tag{3.1}$$

The only properties we will need from these matrices are that D_0 is diagonal, $D_0^*D_0$ has distinct, positive, real diagonal entries, and U is unitary such that UD_0U^* is in G, but we use these particular matrices for concreteness. In Lemma 3.6 we will show that D as in (3.1) is indeed in G.

The group $\mathrm{SU}_{n/2,n/2}$ is not compact, and it may be less familiar than the compact group SU_n, but it is equally useful for our purposes. Both $\mathrm{SU}_{n/2,n/2}$ and SU_n are subgroups of $\mathrm{SL}_n(\mathbb{C})$, and their irreducible representations are exactly the restrictions of those of $\mathrm{SL}_n(\mathbb{C})$. See the appendix of the full version [10] for a brief review of this topic.

The subgroup $\mathrm{U}_n \cap G$ is the block diagonal subgroup $U_{n/2} \times U_{n/2}$, and we need the following subspace, which is close to $\mathrm{Lie}(U_n \cap G)$ but is just a bit smaller:

$$S = \left\{ \left(\begin{array}{c|c} A_0 & 0 \\ \hline 0 & A_1 \end{array} \right) \mid A_0, A_1 \text{ are skew-Hermitian with zeros on the diagonal} \right\}.$$

▶ **Theorem 3.1.** *Let G, D, and S be as above. Let $\mathcal{I}_q \subset \mathbb{C}$ be the set of $a + ib$ with $a, b \in [-\lceil \sqrt{q}/2 \rceil, \lceil \sqrt{q}/2 \rceil]$. Define the following ε-parametrized subsets of G:*

$$\begin{aligned} X &= \{D^{-1} \exp(\varepsilon A) D \mid A \in S\}, \\ Z &= \{D \exp(\varepsilon A) D^{-1} \mid A \in S\}, \\ Y_q &= \big\{ \exp(\varepsilon A) \mid A \in S, \text{ entries of } A \text{ lie in } \mathcal{I}_q \big\}. \end{aligned}$$

Then there exist f_X, f_Z, p_X, p_Z, and p_0 of degree $O(q)$ satisfying the hypothesis of Lemma 2.11. Hence by Lemma 2.11, there exist three ε-parametrized subsets $X_q, Y_q, Z_q \subseteq G$, all of size at least $q^{n^2/4 - n/2}$, which satisfy the TPP and admit border-separating polynomials of degree $O(q)$.

3.2 A precursor construction in $\mathrm{GL}_n(\mathbb{C})$

In preparation for the proof of Theorem 3.1, we first establish a precursor in the containing group $\mathrm{GL}_n(\mathbb{C})$. Recall that the unitary group U_n is the set of matrices M for which $M^*M = I$, and it has half the dimension of the containing group.[6] We will shortly show that three conjugates of U_n in $\mathrm{GL}_n(\mathbb{C})$ almost satisfy the TPP (Theorem 3.3). The proof will make use of the following lemma:

▶ **Lemma 3.2.** *If $M \in \mathrm{U}_n$ and $D = UD_0U^*$ as defined above, then*

$$\mathrm{tr}(D^*M^*D^*DMD) \leq \mathrm{tr}((D^*D)^2)$$

with equality if and only if U^*MU is diagonal.

For a proof, see the full version of the paper [10].

This lemma gives a convenient way to prove that the three subgroups defined in the following theorem satisfy the TPP up to a "failure at the diagonal":

▶ **Theorem 3.3.** *Let $D = UD_0U^*$ be as defined as above. Then the three subgroups $X = D^{-1}\mathrm{U}_nD$, $Y = \mathrm{U}_n$, and $Z = D\mathrm{U}_nD^{-1}$ of $\mathrm{GL}_n(\mathbb{C})$ satisfy the following property. For all*

$$x = D^{-1}M_1D \in X, \quad y = M_2 \in Y, \quad z = DM_3D^{-1} \in Z,$$

if $xyz = I$, then U^*M_1U, U^*M_2U, and U^*M_3U are diagonal matrices.

For a proof, see the full version of the paper [10].

In the variant presented in the next section, we will form our TPP sets as exponentials of Lie algebra elements, and for this, we need a version of Lemma 3.2 that describes the deviation of $\mathrm{tr}(D^*M^*D^*DMD)$ below $\mathrm{tr}((D^*D)^2)$ explicitly as a function of the infinitesimal. Recall that the Lie algebra of U_n is the set of skew-Hermitian matrices.

[6] The unitary group is a real algebraic group, so we need to count real dimensions. The containing group has $2n^2$ real dimensions and the unitary group has n^2 real dimensions.

▶ **Lemma 3.4.** *Let $D = UD_0U^*$ be as defined above, with $d_1, d_2, \ldots, d_n$ being the diagonal entries of D_0. Let A, B be skew-Hermitian matrices and let $M = \exp(\varepsilon A)(\exp(\varepsilon B))^{-1}$. Then*

$$\operatorname{tr}(D^*M^*D^*DMD) = \operatorname{tr}((D^*D)^2) - \varepsilon^2 \sum_{i<j} (|d_i|^2 - |d_j|^2)^2 |C[i,j]|^2 + O(\varepsilon^3),$$

where $C = U^*(A - B)U$. In particular, this quantity is $\operatorname{tr}((D^*D)^2) + O(\varepsilon^3)$ if and only if C is diagonal.

For a proof, see the full version of the paper [10].

In preparation for the next subsection, we now describe some invariant polynomials in $\mathrm{Inv}_{X,Z}$ that we'll use below, where $X = D^{-1}\mathrm{U}_n D$ and $Z = D\mathrm{U}_n D^{-1}$. For an $n \times n$ matrix M and subsets $S, T \subseteq [n]$, we use $M_{S,T}$ to denote the $|S| \times |T|$ sub-matrix of M whose rows are indexed by the elements of S and whose columns are indexed by the elements of T.

▶ **Lemma 3.5.** *For each k, the function $p_k(M)$ defined as*

$$\sum_{\substack{S,T \subseteq [n] \\ |S|=|T|=k}} |\det((DMD)_{S,T})|^2$$

lies in $\mathrm{Inv}_{X,Z}$, where $X = D^{-1}\mathrm{U}_n D$ and $Z = D\mathrm{U}_n D^{-1}$.

We have been implicitly using this fact for p_1, which happens to be the complex Frobenius norm-squared function. In particular, $\operatorname{tr}(D^*M^*D^*DMD) = p_1(M)$. For a proof of Lemma 3.5, see the full version of the paper [10].

Note that because of the complex norm, these functions are not polynomials in the complex matrix entries (but they are polynomials in the entries of the natural real embedding into twice the dimension). In the next section we restrict the containing group to a special unitary group, which has the crucial side effect of making these functions polynomials in the *complex* matrix entries, because the complex conjugate of a given entry can be found as low-degree polynomial of the entries of the inverse matrix.

3.3 Proof of Theorem 3.1

Recall that the containing group of our construction in Theorem 3.1 is the following unitary group:

$$G = \mathrm{SU}_{n/2,n/2} = \{M \in \mathrm{GL}_n(\mathbb{C}) \mid M^*QM = Q \text{ and } \det(M) = 1\},$$
$$\text{with } Q = \left(\begin{array}{c|c} I & \\ \hline & -I \end{array}\right).$$

Our TPP triple in G stated in Theorem 3.1 is obtained by taking the "almost TPP" triple of Theorem 3.3 (recall that the construction almost satisfies the TPP except for a "failure at the diagonal") and intersecting it with G. In this section, we will fix the failure at the diagonal using Lemma 2.11, to get a genuine TPP construction and border-separating polynomials of the optimal degree.

Recall that our construction in G is given by

$$X = \{D^{-1}\exp(\varepsilon A)D \mid A \in S\},$$
$$Z = \{D\exp(\varepsilon A)D^{-1} \mid A \in S\},$$
$$Y_q = \left\{\exp(\varepsilon A) \mid A \in S, \text{ entries of } A \text{ lie in } \mathcal{I}_q\right\},$$

where $\mathcal{I}_q \subset \mathbb{C}$ is the set of $a + ib$ with $a, b \in [-\lceil\sqrt{q}/2\rceil, \lceil\sqrt{q}/2\rceil]$ and

$$S = \left\{ \left(\begin{array}{c|c} A_0 & 0 \\ \hline 0 & A_1 \end{array} \right) \mid A_0, A_1 \text{ are skew-Hermitian with zeros on the diagonal} \right\}.$$

We will prove Theorem 3.1 in three steps.

Step 1 summary: We show that the sets X and Z in the theorem statement are subsets of $D^{-1}\mathrm{U}_n D \cap G$ and $D\mathrm{U}_n D^{-1} \cap G$ and hence the ring of invariant polynomials contains the polynomials specified in Lemma 3.5. To do this we'll describe the Lie algebra of $U_n \cap G$, and show that D is in G.

Step 2 summary: Next, we describe the separating polynomial for Y_q in this invariant ring, and show that it has degree $O(q)$ as promised.

Step 3 summary: Finally we describe the functions f_X, f_Z, p_X, and p_Z and apply Lemma 2.11. For this we prove a version of the double product property for X and Z.

Step 1: Our TPP triple is defined relative to D, and we now show that the D defined in Section 3.1 is indeed in G as claimed. Recall that we use J for the matrix with ones on the antidiagonal $(n + 1 - i, i)$ and zeroes elsewhere (not the all-ones matrix).

▶ **Lemma 3.6.** *The matrix $D = UD_0U^*$ is an element of G, where*

$$U = \frac{1}{\sqrt{2}} \cdot \left(\begin{array}{c|c} I & J \\ \hline J & -I \end{array} \right),$$

and $D_0 = \mathrm{diag}(n, n-1, n-2, \ldots, n/2+1, (n/2+1)^{-1}, (n/2+2)^{-1}, \ldots, n^{-1})$.

See the full version for the proof [10].

Given that we will define our TPP triple by intersecting the three subgroups from the previous subsection with G, we start by describing the intersection of $Y = \mathrm{U}_n$ with G:

▶ **Lemma 3.7.**

$$\mathrm{U}_n \cap G = \left\{ \left(\begin{array}{c|c} M_{1,1} & 0 \\ \hline 0 & M_{2,2} \end{array} \right) : M_{1,1}, M_{2,2} \in \mathrm{U}_{n/2} \text{ and } \det(M_{1,1}M_{2,2}) = 1 \right\}.$$

See the full version for the proof [10].

A simple corollary, using the fact that the Lie algebra of the unitary group is the algebra of skew-Hermitian matrices, is the following.

▶ **Corollary 3.8.** *The Lie algebra of* $\mathrm{U}_n \cap G$ *is*

$$\mathrm{Lie}(\mathrm{U}_n \cap G) = \left\{ \left(\begin{array}{c|c} A_0 & 0 \\ \hline 0 & A_1 \end{array} \right) \mid A_0^* = -A_0,\ A_1^* = -A_1,\ \mathrm{tr}(A_0) + \mathrm{tr}(A_1) = 0 \right\}.$$

In particular, for any such matrix A and for all sufficiently small $\varepsilon > 0$, the exponential $\exp(\varepsilon A)$ *is in* $\mathrm{U}_n \cap G$.

Now our subsets X and Z are contained in subgroups $D^{-1}\mathrm{U}_n D$ and $D\mathrm{U}_n D^{-1}$, respectively, so the invariant polynomials of Lemma 3.5 are invariant under left multiplication by X and right multiplication by Z as before.

Step 2: A key property of G as the containing group, given that the invariant functions of Lemma 3.5 depend on the complex conjugate of matrix entries in addition to the matrix entries themselves, is described in the following straightforward lemma:

▶ **Lemma 3.9.** *For any $M \in G$,*

$$\overline{M[i,j]} = (-1)^{i+j} \det((QMQ)_{-i,-j})$$

where $(QMQ)_{-i,-j}$ denotes the submatrix obtained by deleting the i-th row and j-th column.

▶ **Theorem 3.10.** *Let D be as in Lemma 3.6 and Y_q as above. For $A, B \in Y_q$, if we set $M = \exp(\varepsilon A)\exp(\varepsilon B)^{-1}$, then*

$$\frac{\operatorname{tr}((D^*D)^2) - \operatorname{tr}(D^*M^*D^*DMD)}{\varepsilon^2} = c + O(\varepsilon),$$

where $2(n!)^2 c$ is a nonnegative integer satisfying $c = O(q)$ and for which $c = 0$ iff $A = B$.

For a proof, see the full version of the paper [10].

We note that $|Y_q| \geq q^{n^2/4-n/2}$ and so if we define the polynomial $r(z)$ to be 1 on 0 and 0 on any other positive integer multiple of $1/(2(n!)^2)$ up to $O(q)$, then the function

$$p_0(M) = r((\operatorname{tr}((D^*D)^2) - \operatorname{tr}(D^*M^*D^*DMD))/\varepsilon^2)$$

is a separating function in the ring of invariant functions, of the desired degree $O(q)$. The fact that this is a *polynomial* follows from Lemma 3.9. Because $\operatorname{tr}(D^*M^*D^*DMD)$ is the polynomial $p_1(M)$ of Lemma 3.5, it is invariant under the left/right action of $D^{-1}U_nD$ and DU_nD^{-1}. From step 1 we know that $X \subseteq D^{-1}U_nD$ and $Z \subseteq DU_nD^{-1}$, and hence $p_0 \in \mathrm{Inv}_{X,Z}$.

Step 3: We now turn to describing the functions f_X, f_Z, p_X, and p_Z, beginning with showing that X and Z satisfy a version of the double product property.

▶ **Lemma 3.11.** *For $x \in D^{-1}SD$ and $y \in DSD^{-1}$, $x + y = 0$ iff $x = y = 0$.*

For a proof, see the full version of the paper [10].

▶ **Lemma 3.12.** *For X, Z as above, there exist functions f_X, f_Z and polynomials p_X, p_Z over $\mathbb{C}$ satisfying the first condition of Lemma 2.11 with $d_X = d_Z = n^2/4 - n/2$.*

For a proof, see the full version of the paper [10].

This completes the proof of Theorem 3.1.

4 Conclusions and open problems

In this paper we gave an extension of the group–theoretic framework of [14] to infinite groups (Theorem 2.2). Within this framework we explored constructions in Lie groups, raising the key question: do there exist three subsets in GL_n satisfying the TPP, of size at least $q^{n^2/2-o_n(n)}$, and admitting separating polynomials of degree at most $q^{1+o_q(1)}$? If the answer is yes, then $\omega = 2$ (Corollary 2.8).

Towards obtaining such a construction, we developed tools using invariant theory and Lie algebras to simplify the task of designing separating polynomials (Subsection 2.4). We then put these tools to use in Section 3 to obtain a construction in U_n satisfying the target degree bound, with sets of size q^m with m approaching half the ambient dimension.

This raises several directions for future research:

- Can one obtain a construction with sets of size q^m for m approaching half the ambient dimension and separating polynomials of degree at most $q^{1+o_q(q)}$, but in GL_n rather than U_n?

- Can one obtain a construction with separating polynomials of degree at most $q^{1+o_q(q)}$ in $G_n = \mathrm{GL}_n$ or $G_n = \mathrm{U}_n$ with sets of size $q^{\dim G_n/2-o(n)}$, rather than $q^{\dim G_n/2-\Theta(n)}$?
- Our general framework opens up the possibility of using other infinite groups (not necessarily Lie groups), which remains to be explored.
- Another type of construction that is now possible is one in a single, fixed infinite group (say, GL_3), with growing families of sets (X_q, Y_q, Z_q). In contrast, our current constructions require us to take n growing as well, or more generally, a growing family of containing groups.

References

1. Josh Alman and Virginia Vassilevska Williams. Limits on all known (and some unknown) approaches to matrix multiplication. *SIAM Journal on Computing*, 52(6):FOCS18-285–FOCS18-315, 2023. doi:10.1137/19M124695X.
2. Josh Alman and Virginia Vassilevska Williams. A refined laser method and faster matrix multiplication. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 522–539. SIAM, 2021. doi:10.1137/1.9781611976465.32.
3. Andris Ambainis, Yuval Filmus, and François Le Gall. Fast matrix multiplication: Limitations of the Coppersmith–Winograd method. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing (STOC 2015)*, pages 585–593. ACM, 2015. doi:10.1145/2746539.2746554.
4. Andrew Baker. *Matrix groups: an introduction to Lie group theory.* Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, 2002. doi:10.1007/978-1-4471-0183-3.
5. D. Bini. Relations between exact and approximate bilinear algorithms. Applications. *Calcolo*, 17(1):87–97, 1980. doi:10.1007/BF02575865.
6. Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. $O(n^{2.7799})$ complexity for $n \times n$ approximate matrix multiplication. *Information Processing Letters*, 8(5):234–235, 1979. doi:10.1016/0020-0190(79)90113-3.
7. Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A. Grochow, Eric Naslund, William F. Sawin, and Chris Umans. On cap sets and the group-theoretic approach to matrix multiplication. *Discrete Analysis*, 2017:3, 2017. doi:10.19086/da.1245.
8. Jonah Blasiak, Thomas Church, Henry Cohn, Joshua A. Grochow, and Chris Umans. Which groups are amenable to proving exponent two for matrix multiplication?, 2017. Preprint. arXiv:1712.02302.
9. Jonah Blasiak, Henry Cohn, Joshua A. Grochow, Kevin Pratt, and Chris Umans. Matrix multiplication via matrix groups. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.ITCS.2023.19.
10. Jonah Blasiak, Henry Cohn, Joshua A. Grochow, Kevin Pratt, and Chris Umans. Finite matrix multiplication algorithms from infinite groups, 2024. Preprint. arXiv:2410.14905.
11. Matthias Christandl, François Le Gall, Vladimir Lysikov, and Jeroen Zuiddam. Barriers for rectangular matrix multiplication, 2020. Preprint. arXiv:2003.03019.
12. Matthias Christandl, Péter Vrana, and Jeroen Zuiddam. Barriers for fast matrix multiplication from irreversibility. *Theory of Computing*, 17:Paper No. 2, 32, 2021. doi:10.4086/toc.2021.v017a002.
13. Henry Cohn, Robert Kleinberg, Balázs Szegedy, and Christopher Umans. Group-theoretic algorithms for matrix multiplication. In *Proceedings of the 46th Annual Symposium on Foundations of Computer Science (FOCS 2005)*, pages 379–388. IEEE Computer Society, 2005. doi:10.1109/SFCS.2005.39.

14 Henry Cohn and Christopher Umans. A group-theoretic approach to fast matrix multiplication. In *Proceedings of the 44th Annual Symposium on Foundations of Computer Science (FOCS 2003)*, pages 438–449. IEEE Computer Society, 2003. doi:10.1109/SFCS.2003.1238217.

15 Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, 9(3):251–280, 1990. doi:10.1016/S0747-7171(08)80013-2.

16 A. M. Davie and A. J. Stothers. Improved bound for complexity of matrix multiplication. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 143(2):351–369, 2013. doi:10.1017/S0308210511001648.

17 Igor Dolgachev. *Lectures on invariant theory*, volume 296 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2003. doi:10.1017/CBO9780511615436.

18 Ran Duan, Hongxun Wu, and Renfei Zhou. Faster matrix multiplication via asymmetric hashing. In *Proceedings of the 64th Annual Symposium on Foundations of Computer Science (FOCS 2023)*, pages 2129–2138. IEEE Computer Society, 2023. doi:10.1109/FOCS57990.2023.00130.

19 William Fulton and Joe Harris. *Representation theory: a first course*, volume 129 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. doi:10.1007/978-1-4612-0979-9.

20 Roe Goodman and Nolan R. Wallach. *Symmetry, representations, and invariants*, volume 255 of *Graduate Texts in Mathematics*. Springer, Dordrecht, 2009. doi:10.1007/978-0-387-79852-3.

21 François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation (ISSAC 2014)*, pages 296–303. ACM, 2014. doi:10.1145/2608628.2608664.

22 Harriet Pollatsek. *Lie groups: a problem-oriented introduction via matrix groups*. MAA Textbooks. Mathematical Association of America, Washington, DC, 2009.

23 Claudio Procesi. *Lie groups: an approach through invariants and representations*. Universitext. Springer, New York, 2007. doi:10.1007/978-0-387-28929-8.

24 A. Schönhage. Partial and total matrix multiplication. *SIAM Journal on Computing*, 10(3):434–455, 1981. doi:10.1137/0210032.

25 József Solymosi and Van H. Vu. Near optimal bounds for the Erdős distinct distances problem in high dimensions. *Combinatorica*, 28(1):113–125, 2008. doi:10.1007/s00493-008-2099-1.

26 Volker Strassen. Gaussian elimination is not optimal. *Numerische Mathematik*, 13(4):354–356, 1969. doi:10.1007/BF02165411.

27 Volker Strassen. The asymptotic spectrum of tensors. *Journal für die Reine und Angewandte Mathematik*, 384:102–152, 1988. doi:10.1515/crll.1988.384.102.

28 Virginia Vassilevska Williams. Multiplying matrices faster than Coppersmith-Winograd. In *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing (STOC 2012)*, pages 887–898. ACM, 2012. doi:10.1145/2213977.2214056.

29 Virginia Vassilevska Williams, Yinzhan Xu, Zixuan Xu, and Renfei Zhou. New bounds for matrix multiplication: from alpha to omega. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3792–3835. SIAM, 2024. doi:10.1137/1.9781611977912.134.