

An arithmetic-geometric mean inequality for products of three matrices

Arie Israel*, Felix Krahmer† and Rachel Ward‡

June 22, 2015

Abstract

Consider the following noncommutative arithmetic-geometric mean inequality: Given positive-semidefinite matrices $\mathbf{A}_1, \dots, \mathbf{A}_n$, the following holds for each integer $m \leq n$:

$$\frac{1}{n^m} \sum_{j_1, j_2, \dots, j_m = 1}^n \|\mathbf{A}_{j_1} \mathbf{A}_{j_2} \dots \mathbf{A}_{j_m}\| \geq \frac{(n-m)!}{n!} \sum_{\substack{j_1, j_2, \dots, j_m = 1 \\ \text{all distinct}}}^n \|\mathbf{A}_{j_1} \mathbf{A}_{j_2} \dots \mathbf{A}_{j_m}\|,$$

where $\|\cdot\|$ denotes a unitarily invariant norm, including the operator norm and Schatten p -norms as special cases. While this inequality in full generality remains a conjecture, we prove that the inequality holds for products of up to three matrices, $m \leq 3$. The proofs for $m = 1, 2$ are straightforward; to derive the proof for $m = 3$, we appeal to a variant of the classic Araki-Lieb-Thirring inequality for permutations of matrix products.

1 Introduction

The arithmetic-geometric mean (AMGM) inequality says that for any sequence of n non-negative real numbers x_1, x_2, \dots, x_n , the arithmetic mean is greater than or equal to the geometric mean:

$$\frac{x_1 + x_2 + \dots + x_n}{n} \geq (x_1 x_2 \dots x_n)^{1/n}.$$

This can be viewed as a special case ($m = n$) of Maclaurin's inequality:

2010 *Mathematics subject classification.* 15A42

Key words and phrases. Arithmetic-geometric mean inequality, linear algebra, norm inequalities.

*A. Israel is with the Mathematics Department at the University of Texas at Austin, arie@math.utexas.edu

†F. Krahmer is with the Department of Mathematics, Unit M15 Applied Numerical Analysis, Technische Universität München, felix.krahmer@tum.de

‡R. Ward is with the Mathematics Department at the University of Texas at Austin, rward@math.utexas.edu

Proposition 1.1. *If x_1, \dots, x_n are positive scalars and $m \geq n$ then it holds that*

$$\frac{1}{n^m} \sum_{j_1, j_2, \dots, j_m = 1}^n x_{j_1} x_{j_2} \dots x_{j_m} \geq \frac{1}{\binom{n}{m}} \sum_{\substack{\Lambda \subset [n]; \\ |\Lambda| = m}} x_{j_1} x_{j_2} \dots x_{j_m}.$$

See [8] for more details. With a slight abuse of notation, we will refer to both of the above inequalities as AMGM inequalities.

Several noncommutative extensions of Proposition 1.1 have been proven for inequalities involving the product of two matrices. For an overview of these results, we refer the reader to [5]. These inequalities are often stated for a general unitarily invariant (UI) norm. Recall that a norm $\|\cdot\|$ on $M(d)$, the space of complex $d \times d$ matrices, is said to be unitarily invariant if for all $\mathbf{X}, \mathbf{U} \in M(d)$ with \mathbf{U} unitary, one has $\|\mathbf{U}\mathbf{X}\| = \|\mathbf{X}\mathbf{U}\| = \|\mathbf{X}\|$. Examples of UI norms are the Schatten p -norms (including the operator norm and the Hilbert-Schmidt norm) and the Ky Fan k -norms. More generally, every UI norm is a symmetric gauge function of the singular values [2]. The first AMGM inequality for products of two matrices appeared in [4]: If \mathbf{A} and \mathbf{B} are compact operators on a separable Hilbert space, then

$$2\|\mathbf{A}^* \mathbf{B}\| \leq \|\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*\|.$$

The paper [3] extended this result, showing that for arbitrary $d \times d$ matrices $\mathbf{A}, \mathbf{B}, \mathbf{X}$, and for every unitarily invariant norm,

$$2\|\mathbf{A}^* \mathbf{X} \mathbf{B}\| \leq \|\mathbf{A}\mathbf{A}^* \mathbf{X} + \mathbf{X} \mathbf{B}\mathbf{B}^*\|.$$

Most closely related to the results here, Kosaki [10] showed that for positive-semidefinite matrices \mathbf{A} and \mathbf{B} , and for $1/p + 1/q = 1$,

$$\|\mathbf{A}\mathbf{X}\mathbf{B}\| \leq \frac{1}{p} \|\mathbf{A}^p \mathbf{X}\| + \frac{1}{q} \|\mathbf{X}\mathbf{B}^q\|.$$

In the special case $\mathbf{X} = \text{Id}$ and $p = q = 2$, applying this inequality and averaging with respect to the order of \mathbf{A} and \mathbf{B} , reproduces our result for the case of two matrices.

More recently, certain noncommutative extensions of the AMGM inequality have been posed for products of (complex-valued) positive-semidefinite matrices. The following conjecture was posed by Recht and R e [12]:

Conjecture 1.2. *Let $\mathbf{A}_1, \dots, \mathbf{A}_n$ be positive-semidefinite matrices. Then the following inequality holds for each $m \leq n$:*

$$\left\| \frac{1}{n^m} \sum_{j_1, j_2, \dots, j_m = 1}^n \mathbf{A}_{j_1} \mathbf{A}_{j_2} \dots \mathbf{A}_{j_m} \right\| \geq \left\| \frac{(n-m)!}{n!} \sum_{\substack{j_1, j_2, \dots, j_m = 1; \\ \text{all distinct}}}^n \mathbf{A}_{j_1} \mathbf{A}_{j_2} \dots \mathbf{A}_{j_m} \right\|.$$

Here, $\|\cdot\|$ denotes the standard operator norm.

Note that the case $m = 1$ is trivially true, as both sides of the inequality are equal. The conjecture is easily seen to be true for $n = 2, m = 2$, and in this case something stronger can be said: the symmetrized geometric mean precedes the square of the arithmetic mean in the positive definite order: for any $\mathbf{A}, \mathbf{B} \succeq 0$,

$$\left(\frac{1}{2}\mathbf{A} + \frac{1}{2}\mathbf{B}\right)^2 - \left(\frac{1}{2}\mathbf{A}\mathbf{B} + \frac{1}{2}\mathbf{B}\mathbf{A}\right) \succeq 0.$$

Recht and R e also verify that the conjecture holds for general m and n *in expectation* for several classes of random matrices [12].

Later, Ducci [7] posed a variant of the noncommutative AMGM conjecture where the matrix operator norm appears inside the summation on either side:

Conjecture 1.3. *Suppose that $\mathbf{A}_1, \dots, \mathbf{A}_n$ are positive-semidefinite matrices. The following inequality holds for each $m \leq n$:*

$$\frac{1}{n^m} \sum_{j_1, j_2, \dots, j_m = 1}^n \|\mathbf{A}_{j_1} \mathbf{A}_{j_2} \dots \mathbf{A}_{j_m}\| \geq \frac{(n-m)!}{n!} \sum_{\substack{j_1, j_2, \dots, j_m = 1; \\ \text{all distinct}}}^n \|\mathbf{A}_{j_1} \mathbf{A}_{j_2} \dots \mathbf{A}_{j_m}\|. \quad (1)$$

The case $m = 1$ of Conjecture 1.3 is trivially true, as both sides of the inequality are the same. For $m = 2$, as explained above, the conjecture directly follows, for arbitrary UI norms, from a result of Kosaki [10] (we provide a proof without assuming this result in Section 3). Given the positive results for the cases $m = 1$ and $m = 2$, it is tempting to believe that a simple inductive proof could be used to prove the general case. However, the difficulty in this approach, and with the noncommutative AMGM inequalities in general, is that the product of two positive-semidefinite matrices is not necessarily positive-semidefinite. The main result of this paper, proved in Section 4, is a proof of this conjecture for the case $m = 3$. Again, we prove a more general result which holds for any unitarily-invariant norm:

Theorem 1.4 (AMGM inequality for three matrices). *Suppose that $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n \in M(d)$ are positive-semidefinite. Let $\|\cdot\|$ be a unitarily invariant norm on $M(d)$. Then the following AMGM inequality holds:*

$$\frac{1}{n^3} \sum_{i, j, k = 1}^n \|\mathbf{A}_i \mathbf{A}_j \mathbf{A}_k\| \geq \frac{(n-3)!}{n!} \sum_{\substack{i, j, k = 1 \\ \text{all distinct}}}^n \|\mathbf{A}_i \mathbf{A}_j \mathbf{A}_k\|. \quad (2)$$

Remark 1.5. *Unitarily invariant norms $\|\cdot\|$ can also be defined in the infinite-dimensional setting of operators on Hilbert spaces. The subset of compact operators on a Hilbert space \mathcal{H} having finite norm $\|\cdot\|$ defines a self-adjoint ideal \mathcal{J} , called the norm ideal associated to $\|\cdot\|$. As finite-rank operators are dense in this space, our Theorem 4.3, and hence Theorem 1.4, extends to this more general setting using a limiting argument as presented e.g. in Proposition 2.2 of [6].*

To the best of our knowledge, this result represents the first AMGM inequality for products of three matrices. Our proof uses a variant of the Araki-Lieb-Thirring inequality (Theorem 4.3). This inequality is specific to the product of three operators, and it remains an interesting open question whether the result extends to the case $m \geq 4$.

2 Motivation

One can rephrase Conjecture 1.3 in terms of comparing the expectations of random matrices formed by sampling with replacement vs. without replacement. Consider a random variable of the form $X = \|\mathbf{A}_{j_1}\mathbf{A}_{j_2}\dots\mathbf{A}_{j_m}\|$, along with two different probability distributions over the indices j_ℓ : *with replacement* sampling where each index j_ℓ is drawn uniformly with replacement from the index set $[n]$, and *without replacement* sampling where the indices j_ℓ are drawn sequentially, uniformly and *without* replacement from $[n]$. Then $\mathbb{E}_{wr}X$, the expected value of X corresponding to with-replacement sampling, is equal to the LHS expression of 1, while $\mathbb{E}_{wor}X$, the expected value of X corresponding to without-replacement sampling, is equal to the RHS. From this perspective, the noncommutative AMGM conjectures have interesting implications for stochastic optimization problems. For example, Conjecture 1.3 would imply that the expected convergence rate of without-replacement sampling is faster than that of with-replacement sampling for randomized iterative solvers such as the *Kaczmarz* method [9]. We repeat the following example from [7, 12] for completeness. The *Kaczmarz* method is a simple and fast method for solving overdetermined consistent least squares problems: solve for x_* satisfying $\Phi x_* = y$, where $\Phi \in \mathbb{C}^{n \times d}$ with $n \gg d$. Let φ_i^* denote the i th row of Φ . Then, starting from some initial x_0 , the *Kaczmarz* algorithm iterates the following recursion until convergence:

$$x_k = x_{k-1} + \frac{y_k - \langle \varphi_{i_k}, x_{k-1} \rangle}{\|\varphi_{i_k}\|_2^2} \varphi_{i_k}.$$

Set $\mathbf{A}_i = \text{Id} - \varphi_i^* \varphi_i / \|\varphi_i\|_2^2$. Then since $\langle \varphi_i, x_* \rangle = y_i$, we may express the residual after k steps of the *Kaczmarz* algorithm in terms of the matrix product

$$x_k - x_* = \prod_{j=1}^k \mathbf{A}_{i_j} (x_0 - x_*).$$

The residual error can then be bounded by

$$\|\mathbf{A}_{i_1}\mathbf{A}_{i_2}\dots\mathbf{A}_{i_k}(x_0 - x_*)\|_2 \leq \|\mathbf{A}_{i_1}\mathbf{A}_{i_2}\dots\mathbf{A}_{i_k}\| \|x_0 - x_*\|_2.$$

Nonasymptotic convergence rates for the *Kaczmarz* algorithm have been derived [13, 11] in case each row update i_k is selected according to a random update rule, in particular, independently and identically distributed over $[n]$. At the same time, numerical evidence [12] suggest that the convergence rate can be improved by sampling rows independently *without* replacement. Conjecture 1.3 would provide theoretical justification to these observations, showing that with-replacement sampling cannot outperform without-replacement sampling, in expectation with respect to the draw of the row

indices. Since the randomized Kaczmarz algorithm can be viewed as a special case of stochastic gradient descent [11], similar remarks about with-replacement vs. without replacement sampling could likely be made for stochastic gradient descent algorithms more generally.

3 AMGGM inequality for products of two matrices

As a warm up to the main result, we first prove the analog of Theorem 1.4 for products of two positive-semidefinite matrices. Throughout the section, we fix a unitarily invariant norm $\|\cdot\|$ on $M(d)$. In the proof, we will use several basic results about unitarily invariant norms. First, a variant of Hölder's inequality:

Lemma 3.1 (Exercise IV.2.7 of [2]). *Suppose that $p, q, r > 0$ satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then for all $\mathbf{A}, \mathbf{B} \in M(d)$ it holds that*

$$\|\|\mathbf{AB}\|^r\|^{1/r} \leq \|\|\mathbf{A}\|^p\|^{1/p} \|\|\mathbf{B}\|^q\|^{1/q}.$$

Here, we have used the standard notation $|\mathbf{A}| := (\mathbf{A}^* \mathbf{A})^{1/2}$. It follows from Lemma 3.1 that for all $\mathbf{A}, \mathbf{B} \in M(d)$ one has

$$\|\|\mathbf{AB}\| = \|\|\mathbf{AB}\| \leq \|\|\mathbf{A}^2\|^{1/2} \cdot \|\|\mathbf{B}^2\|^{1/2} \leq \frac{1}{2} \|\|\mathbf{A}^2\| + \frac{1}{2} \|\|\mathbf{B}^2\|, \quad (3)$$

where in the first equality we used the polar decomposition and the unitary invariance of $\|\cdot\|$, and in the last inequality we used the scalar AMGGM inequality. Similarly, by a repeated application of Lemma 3.1, for all $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \in M(d)$ we have

$$\begin{aligned} \|\|\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3\| &\leq \|\|\mathbf{X}_1\|^3\|^{1/3} \cdot \|\|\mathbf{X}_2 \mathbf{X}_3\|^{3/2}\|^{2/3} \leq \|\|\mathbf{X}_1\|^3\|^{1/3} \cdot \|\|\mathbf{X}_2\|^3\|^{1/3} \cdot \|\|\mathbf{X}_3\|^3\|^{1/3} \\ &\leq \frac{1}{3} \|\|\mathbf{X}_1\|^3\| + \frac{1}{3} \|\|\mathbf{X}_2\|^3\| + \frac{1}{3} \|\|\mathbf{X}_3\|^3\|. \end{aligned} \quad (4)$$

Alternatively, if $\mathbf{X}_2 \succeq 0$ then we obtain

$$\|\|\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3\| \leq \frac{1}{2} \|\|\mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_1\| + \frac{1}{2} \|\|\mathbf{X}_3 \mathbf{X}_2 \mathbf{X}_3\| \quad (5)$$

by applying (3) to $\mathbf{A} = \mathbf{X}_1 \mathbf{X}_2^{1/2}$ and $\mathbf{B} = \mathbf{X}_2^{1/2} \mathbf{X}_3$.

Proposition 3.2. *Let $n \geq 2$ and $d \geq 1$, and suppose that $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n \in M(d)$ are positive-semidefinite. Then the following arithmetic-geometric mean inequality holds:*

$$\frac{1}{n^2} \sum_{j, k=1}^n \|\|\mathbf{A}_j \mathbf{A}_k\| \geq \frac{1}{n(n-1)} \sum_{\substack{j, k=1 \\ j \neq k}}^n \|\|\mathbf{A}_j \mathbf{A}_k\|.$$

Proof. Rearranging and canceling like terms, the inequality reduces to

$$(n-1) \sum_{j=1}^n \|\|\mathbf{A}_j^2\| \geq \sum_{\substack{j, k=1 \\ j \neq k}}^n \|\|\mathbf{A}_j \mathbf{A}_k\|,$$

which is equivalent to

$$\sum_{j=1}^n \sum_{k=j+1}^n (\|A_j^2\| + \|A_k^2\| - \|A_j A_k\| - \|A_k A_j\|) \geq 0.$$

To show that each of the summands is nonnegative, we apply (3) both for $\|A_i A_j\|$ and for $\|A_j A_i\|$, obtaining

$$\|A_i^2\| + \|A_j^2\| \geq \|A_i A_j\| + \|A_j A_i\|.$$

This proves the proposition. □

4 AMGM inequality for products of three matrices

Before proving the main result, Theorem 1.4, we will need to establish an ALT-type inequality for unitarily invariant norms. The classic version of this inequality states:

Theorem 4.1 (Araki-Lieb-Thirring [1]). *For all $r \geq 1$ and $q > 0$ one has*

$$\mathrm{Tr}[(B^r A^r B^r)^q] \geq \mathrm{Tr}[(BAB)^{rq}].$$

A similar statement holds for the matrix operator norm:

Lemma 4.2. *For all $r \geq 1$ and $s > 0$ one has*

$$\|(B^r A^r B^r)^s\| \leq \|(BAB)^{rs}\|.$$

Proof. We apply the ALT inequality with $q = st$, for $t > 0$, to obtain

$$\mathrm{Tr}[(B^r A^r B^r)^{st}]^{1/t} \geq \mathrm{Tr}[(BAB)^{rst}]^{1/t}.$$

Take the limit as $t \rightarrow +\infty$. The desired inequality follows as $\lim_{t \rightarrow +\infty} \mathrm{Tr}[\mathbf{X}^t]^{1/t} = \|\mathbf{X}\|$ for any $\mathbf{X} \succeq 0$. □

In fact, we can generalize Lemma 4.2 to any unitarily invariant norm using a standard trick in matrix analysis known as the “antisymmetric tensor power” trick. We defer the proof to the appendix.

Theorem 4.3. *For any unitarily invariant norm $\|\cdot\|$ on $M(d)$, and for any $r \geq 1$ and $s > 0$, the following holds for any two positive-semidefinite matrices \mathbf{A} and \mathbf{B} in $M(d)$:*

$$\|(B^r A^r B^r)^s\| \leq \|(BAB)^{rs}\|.$$

In particular, we will use the following corollary:

Corollary 4.4. *For any two positive-semidefinite matrices $\mathbf{A}, \mathbf{B} \in M(d)$, and for any unitarily invariant norm,*

$$\|\mathbf{B}^2 \mathbf{A}\| \geq \|\mathbf{B} \mathbf{A} \mathbf{B}\|.$$

Proof. Apply Theorem 4.3 with $r = 2$ and $s = 1/2$ to obtain

$$\left\| \sqrt{\mathbf{B}^2 \mathbf{A}^2 \mathbf{B}^2} \right\| \leq \left\| \mathbf{B} \mathbf{A} \mathbf{B} \right\| \quad \text{for any } \mathbf{A}, \mathbf{B} \in M(d), \mathbf{A}, \mathbf{B} \succeq 0.$$

Note that $\sqrt{\mathbf{B}^2 \mathbf{A}^2 \mathbf{B}^2} = \sqrt{(\mathbf{B}^2 \mathbf{A})(\mathbf{B}^2 \mathbf{A})^*} = |\mathbf{B}^2 \mathbf{A}|$. By the unitary invariance of the norm $\|\cdot\|$ we thus obtain that

$$\left\| \mathbf{B}^2 \mathbf{A} \right\| = \left\| |\mathbf{B}^2 \mathbf{A}| \right\| \leq \left\| \mathbf{B} \mathbf{A} \mathbf{B} \right\|,$$

as desired. \square

With Corollary 4.4 in hand, we are now ready to prove the main result, Theorem 1.4.

Proof of Theorem 1.4. Rearranging and canceling like terms, the desired inequality (2) is equivalent to

$$(n-2)(n-1) \sum_{\substack{i, j, k = 1 \\ \text{not all distinct}}}^n \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_k \right\| \geq (3n-2) \sum_{\substack{i, j, k = 1 \\ \text{all distinct}}}^n \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_k \right\|. \quad (6)$$

By appeal to Corollary 4.4, we obtain the bounds

$$\left\| \mathbf{A}_i^2 \mathbf{A}_j \right\| \geq \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_i \right\|.$$

By the unitary invariance of $\|\cdot\|$ we know that $\|\mathbf{X}\| = \|\mathbf{X}^*\|$ for all $\mathbf{X} \in M(d)$. Thus, by the same reasoning as above, we have $\left\| \mathbf{A}_i \mathbf{A}_j^2 \right\| = \left\| \mathbf{A}_j^2 \mathbf{A}_i \right\| \geq \left\| \mathbf{A}_j \mathbf{A}_i \mathbf{A}_j \right\|$. Applying these lower bounds to the LHS of (6) gives:

LHS of (6)

$$\begin{aligned} &= (n-2)(n-1) \left[\sum_{j=1}^n \left\| \mathbf{A}_j^3 \right\| + \sum_{\substack{i, j = 1 \\ i \neq j}}^n (\left\| \mathbf{A}_i^2 \mathbf{A}_j \right\| + \left\| \mathbf{A}_i \mathbf{A}_j^2 \right\|) + \sum_{\substack{i, j = 1 \\ i \neq j}}^n \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_i \right\| \right] \\ &\geq (n-2)(n-1) \sum_{j=1}^n \left\| \mathbf{A}_j^3 \right\| + 3(n-2)(n-1) \sum_{\substack{i, j = 1 \\ i \neq j}}^n \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_i \right\|. \end{aligned} \quad (7)$$

We apply (5) to bound the expressions $\left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_k \right\|$ and $\left\| \mathbf{A}_k \mathbf{A}_j \mathbf{A}_i \right\|$, thus obtaining

$$\left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_i \right\| + \left\| \mathbf{A}_k \mathbf{A}_j \mathbf{A}_k \right\| \geq \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_k \right\| + \left\| \mathbf{A}_k \mathbf{A}_j \mathbf{A}_i \right\|.$$

Summing this estimate over pairwise distinct i, j, k gives

$$2(n-2) \sum_{\substack{i, j = 1 \\ i \neq j}}^n \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_i \right\| \geq 2 \sum_{\substack{i, j, k = 1 \\ \text{all distinct}}}^n \left\| \mathbf{A}_i \mathbf{A}_j \mathbf{A}_k \right\|.$$

Continuing the bound from (7), we have

$$\text{LHS of (6)} \geq (n-2)(n-1) \sum_{j=1}^n \|\mathbf{A}_j^3\| + 3(n-1) \sum_{\substack{i,j,k=1 \\ \text{all distinct}}}^n \|\mathbf{A}_i \mathbf{A}_j \mathbf{A}_k\|.$$

The desired inequality (6) will follow if the RHS above is greater or equal to the RHS of (6). Rearranging and canceling like terms, this is equivalent to asking whether

$$(n-2)(n-1) \sum_{j=1}^n \|\mathbf{A}_j^3\| \geq \sum_{\substack{i,j,k=1 \\ \text{all distinct}}}^n \|\mathbf{A}_i \mathbf{A}_j \mathbf{A}_k\|.$$

But indeed this inequality follows directly from averaging (4) with $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) = (\mathbf{A}_i, \mathbf{A}_j, \mathbf{A}_k)$ over all choices of pairwise distinct i, j, k . This finishes the proof. \square

Discussion

There are certain difficulties that arise in the case $m = 4$ which prevent us from extending the proof technique of Theorem 1.4, even in the setting of matrices. “Loopy” terms such as $\|\mathbf{A}_1 \mathbf{A}_2 \mathbf{A}_3 \mathbf{A}_2\|$ start to appear at $m = 4$, and it is not clear how to pair these terms together to obtain a lower bound of distinct-term products, $\|\mathbf{A}_1 \mathbf{A}_2 \mathbf{A}_3 \mathbf{A}_4\|$. Simple inductive arguments building on the cases $m = 1, 2, 3$ are also difficult as the product of two positive-semidefinite operators is not necessarily positive-semidefinite, and there are not many inequalities concerning products of four positive-semidefinite operators. This remains a very compelling open problem.

Acknowledgments

We are thankful to Afonso Bandeira, Dustin Mixon, Deanna Needell, Ben Recht, and Robert Schaback for helpful conversations on this topic. Kraemer was supported by the German Science Foundation (DFG) in the context of the Emmy Noether Junior Research Group KR 4512/1-1 (RaSenQuaSI). Ward was partially supported by an NSF CAREER award, DOD-Navy grant N00014-12-1-0743, and an AFOSR Young Investigator Program award.

References

- [1] H. Araki. On an inequality of Lieb and Thirring. *Letters in Mathematical Physics*, 19(2):167–170, 1990.
- [2] R. Bhatia. Matrix analysis. *Springer*, 1997.
- [3] R. Bhatia and C. Davis. More matrix forms of the arithmetic-geometric mean inequality. *Siam. J. Matrix Anal. Appl.*, 14(1):132–136, 1993.

- [4] R. Bhatia and F. Kittaneh. On the singular values of a product of operators. *Siam J. Matrix Anal. Appl.*, 11:272–277, 1990.
- [5] R. Bhatia and F. Kittaneh. The matrix arithmetic-geometric mean inequality revisited. *Linear Algebra Appl.*, 428(8-9):2177–2191, 2008.
- [6] C. Cano, I. Mosconi, and D. Stojanoff. Some operator inequalities for unitarily invariant norms. *Revista de la Union Matematica Argentina*, 46(1):53–66, 2005.
- [7] J. Ducci. Commentary on “Towards a noncommutative arithmetic-geometric mean inequality” by B. Recht and C. Ré. *JMLR: Workshop and conference proceedings*, 23, 2012.
- [8] G. Hardy, J. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 2nd edition, 1952.
- [9] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *International Bulletin of the Polish Academy of Sciences, Letters A*, pages 335–357, 1937.
- [10] H. Kosaki. Arithmetic-geometric mean and related inequalities for operators. *J. Func. Anal.*, 156:429–451, 1998.
- [11] D. Needell, N. Srebro, and R. Ward. Stochastic gradient descent and the randomized Kaczmarz algorithm. *arXiv preprint arXiv:1310.5715*, 2013.
- [12] B. Recht and C. Ré. Beneath the valley of the noncommutative arithmetic-geometric mean inequality: conjectures, case-studies, and consequences. *arXiv preprint arXiv:1202.4184*, 2012.
- [13] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15(2):262–278, 2009.

5 Appendix: Proof of Theorem 4.3

Given $\mathbf{A}, \mathbf{B} \in M(d)$ with $\mathbf{A}, \mathbf{B} \succeq 0$, and given $r \geq 1$ and $s > 0$, our aim in this section is to verify the ALT-type inequality $\|(\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s\| \leq \|(\mathbf{B} \mathbf{A} \mathbf{B})^{r+s}\|$ for a general unitarily invariant norm $\|\cdot\|$ on $M(d)$. The proof uses techniques from matrix analysis and we will make frequent use of results in [2]. For the rest of this section, all matrices are assumed to be positive-semidefinite.

5.1 Antisymmetric tensorization

In what follows, we write $\lambda_1(\mathbf{X}) \geq \lambda_2(\mathbf{X}) \geq \lambda_3(\mathbf{X}) \geq \dots$, to denote the sequence of eigenvalues of a matrix \mathbf{X} , written in non-increasing arrangement.

For $1 \leq k \leq d$, we denote the space $\mathcal{H}_k = \wedge^k \mathbb{R}^d$ for the order- k antisymmetric tensor power of \mathbb{R}^d , which consists of all formal real linear combinations of symbols of the form

$$x_1 \wedge x_2 \wedge \dots \wedge x_k,$$

where $x_1, \dots, x_k \in \mathbb{R}^d$. We regard \mathcal{H}_k as a vector space in such a way that addition and scalar multiplication are multilinear in the variables (x_1, \dots, x_k) . Namely,

$$\begin{cases} x_1 \wedge \dots \wedge x_l \wedge \dots \wedge x_k + x_1 \wedge \dots \wedge x'_l \wedge \dots \wedge x_k = x_1 \wedge \dots \wedge (x_l + x'_l) \wedge \dots \wedge x_k. \\ c \cdot (x_1 \wedge \dots \wedge x_l \wedge \dots \wedge x_k) = x_1 \wedge \dots \wedge (cx_l) \wedge \dots \wedge x_k. \end{cases}$$

We also require that the wedge product \wedge is antisymmetric with respect to coordinate interchanges. Namely, for all $1 \leq i < j \leq k$ we have

$$x_1 \wedge \dots \wedge x_i \wedge \dots \wedge x_j \wedge \dots \wedge x_k = -x_1 \wedge \dots \wedge x_j \wedge \dots \wedge x_i \wedge \dots \wedge x_k.$$

From this definition one sees that the dimension of \mathcal{H}_k is $\binom{d}{k}$.

Given a sequence of matrices $\mathbf{A}_1, \dots, \mathbf{A}_k \in M(d)$ – which we regard as linear operators on \mathbb{R}^d – we form the antisymmetric tensor matrix $\mathbf{A}_1 \wedge \dots \wedge \mathbf{A}_k$, which represents a linear operator on the tensor space \mathcal{H}_k defined by

$$(\mathbf{A}_1 \wedge \dots \wedge \mathbf{A}_k)(x_1 \wedge \dots \wedge x_k) = (\mathbf{A}_1 x_1) \wedge \dots \wedge (\mathbf{A}_k x_k) \quad \text{for } x_1 \wedge \dots \wedge x_k \in \mathcal{H}_k.$$

Given a matrix $\mathbf{A} \in M(d)$, we denote

$$\wedge^k \mathbf{A} := \overbrace{\mathbf{A} \wedge \mathbf{A} \wedge \dots \wedge \mathbf{A}}^{k \text{ times}}, \quad \text{a linear operator on } \mathcal{H}_k.$$

We present a few basic properties of antisymmetric tensorization. Proofs of properties 1-5, along with a further discussion, are found in Section I.5 of [2].

1. $(\wedge^k \mathbf{A})(\wedge^k \mathbf{B}) = \wedge^k(\mathbf{A}\mathbf{B})$.
2. $(\wedge^k \mathbf{A})^* = \wedge^k \mathbf{A}^*$.
3. $(\wedge^k \mathbf{A})^{-1} = \wedge^k \mathbf{A}^{-1}$ if \mathbf{A} is invertible.
4. If \mathbf{A} is unitary or positive-semidefinite then so is $\wedge^k \mathbf{A}$.
5. If $\lambda_{i_1}, \lambda_{i_2}, \dots, \lambda_{i_k}$ are eigenvalues of \mathbf{A} associated to linearly independent eigenvectors $u_{i_1}, u_{i_2}, \dots, u_{i_k}$, and if $i_1 > i_2 > \dots > i_k$, then $\lambda_{i_1} \cdot \lambda_{i_2} \cdot \dots \cdot \lambda_{i_k}$ is an eigenvalue of $\wedge^k \mathbf{A}$ associated to the eigenvector $u_{i_1} \wedge u_{i_2} \wedge \dots \wedge u_{i_k}$. In particular, if \mathbf{A} is diagonalizable with eigenvalues λ_j , $1 \leq j \leq d$, then $\wedge^k \mathbf{A}$ is diagonalizable with eigenvalues $\lambda_{i_1} \cdot \dots \cdot \lambda_{i_k}$, $1 \leq i_1 < \dots < i_k \leq d$.
6. If \mathbf{A} is positive-semidefinite then $(\wedge^k \mathbf{A})^s = \wedge^k \mathbf{A}^s$ for all $s > 0$.

Proof of property 6: Fix a complete set of eigenvalues $\lambda_1 \geq \dots \geq \lambda_d \geq 0$ and eigenvectors u_1, \dots, u_d for \mathbf{A} . The eigenvalues of \mathbf{A}^s are given by $\lambda_1^s, \dots, \lambda_d^s$ and the eigenvectors are unchanged. According to property 5 we know that the eigenvalues/eigenvectors for $\wedge^k \mathbf{A}^s$ are given by $\lambda_{i_1}^s \cdot \dots \cdot \lambda_{i_k}^s$ and $u_{i_1} \wedge \dots \wedge u_{i_k}$, where $1 \leq i_1 < \dots < i_k \leq d$. On the other hand, by property 5 we know the eigenvalues of $\wedge^k \mathbf{A}$ are $\lambda_{i_1} \cdot \dots \cdot \lambda_{i_k}$ associated to the eigenvectors $u_{i_1} \wedge \dots \wedge u_{i_k}$; hence the eigenvalues of $(\wedge^k \mathbf{A})^s$ are $\lambda_{i_1}^s \cdot \dots \cdot \lambda_{i_k}^s$ associated to the same eigenvectors. Therefore, it holds that $(\wedge^k \mathbf{A})^s = \wedge^k \mathbf{A}^s$.

Lemma 5.1. *For all $r \geq 1$ and $s > 0$ one has*

$$\prod_{i=1}^k \lambda_i((\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s) \leq \prod_{i=1}^k \lambda_i(\mathbf{B} \mathbf{A} \mathbf{B})^{rs} \quad \text{for all } k = 1, 2, \dots, d.$$

Proof. We apply the norm inequality in Lemma 4.2 to the tensor matrices $\wedge^k \mathbf{A}$ and $\wedge^k \mathbf{B}$. Thus, $\|((\wedge^k \mathbf{B})^r (\wedge^k \mathbf{A})^r (\wedge^k \mathbf{B})^r)^s\| \leq \|((\wedge^k \mathbf{B}) \cdot (\wedge^k \mathbf{A}) \cdot (\wedge^k \mathbf{B}))^{rs}\|$. Applying subsequently properties 1 and 6, we obtain $\|\wedge^k (\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s\| \leq \|\wedge^k (\mathbf{B} \mathbf{A} \mathbf{B})^{rs}\|$. Recall that the eigenvalues of $(\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s$ are ordered in non-increasing arrangement: $\lambda_1((\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s) \geq \lambda_2((\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s) \geq \dots \geq \lambda_d((\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s)$. Therefore, due to property 5, the operator norm of $\wedge^k (\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s$, which is the largest eigenvalue of $\wedge^k (\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s$ is

$$\lambda_1((\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s) \cdots \lambda_k((\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s).$$

By similar reasoning, the operator norm of $\wedge^k (\mathbf{B} \mathbf{A} \mathbf{B})^{rs}$ is

$$\lambda_1((\mathbf{B} \mathbf{A} \mathbf{B})^{rs}) \cdots \lambda_k((\mathbf{B} \mathbf{A} \mathbf{B})^{rs}).$$

This establishes the desired inequality. \square

5.2 Weak Majorization

Recall that $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d) \in \mathbb{R}^d$ is *weakly majorized* by $\beta = (\beta_1, \beta_2, \dots, \beta_d) \in \mathbb{R}^d$ – written in shorthand as $\alpha \prec_w \beta$ – if for all $k = 1, 2, \dots, d$ it holds that

$$\sum_{i=1}^k \alpha_i \leq \sum_{i=1}^k \beta_i.$$

We borrow the following lemma from [2]; see Exercise II.3.5.

Lemma 5.2. *Assume that $x = (x_i)_{i=1}^d$ and $y = (y_i)_{i=1}^d$ satisfy $x_1 \geq x_2 \geq \dots \geq x_d \geq 0$ and $y_1 \geq y_2 \geq \dots \geq y_d \geq 0$. If additionally*

$$\prod_{i=1}^k x_i \leq \prod_{i=1}^k y_i \quad \text{for all } k = 1, \dots, d,$$

then $x \prec_w y$.

From Lemma 5.1 and Lemma 5.2 we obtain a weak-majorization inequality between the eigenvalues of the matrices of interest. Namely, for any $r \geq 1$ and $s > 0$ we have

$$(\lambda_j((\mathbf{B}^r \mathbf{A}^r \mathbf{B}^r)^s))_{j=1}^d \prec_w (\lambda_j(\mathbf{B} \mathbf{A} \mathbf{B})^{rs})_{j=1}^d. \quad (8)$$

By the Fan Dominance Theorem (see Theorem IV.2.2 in [2]), the weak majorization in (8) implies the desired estimate, Theorem 4.3.