ON THE SYMMETRIZED ARITHMETIC-GEOMETRIC MEAN INEQUALITY FOR OPERATORS

WAFAA ALBAR, MARIUS JUNGE+, AND MINGYU ZHAO

We study the symmetrized noncommutative arithmetic geometric mean inequality introduced by Recht and Ré in [10]
\[ \left\| \frac{(n-d)!}{n!} \sum_{j_1, \ldots, j_d \text{ different}} A_{j_1}^* A_{j_2} \ldots A_{j_d} A_{j_d} \ldots A_{j_2} A_{j_1} \right\| \leq C(d, n) \left\| \frac{1}{n} \sum_{j=1}^{n} A_{j}^* A_{j} \right\|^d. \]

Complementing the results from [10], we find upper bounds for $C(d,n)$ under additional assumptions. Moreover, using free probability, we show that $C(d,n) > 1$, thereby disproving the most optimistic conjecture from [10]. We also prove a deviation result for the symmetrized-AGM inequality which shows that the symmetric inequality almost holds for many classes of random matrices. Finally we apply our results to the incremental gradient method (IGM).

1. INTRODUCTION

The arithmetic-geometric mean (AGM) inequality is of fundamental importance in mathematical analysis. Augustin Cauchy (1789-1857) was the first to prove AGM in 1821. Then Liouville, Hurwitz, Steffensen, Bohr, Riesz, Sturm, Rado, Hardy, Littlewood, and Polya offered alternative proofs of the AGM inequalities in the same year. Much more recently, Ré and Recht in [10] realize that sampling without-replacement outperforms sampling with replacement provided a noncommutative version of the arithmetic-geometric mean (AGM) inequality holds. They formulate several conjectures on AGM with connection to machine learning such as the incremental gradient method, empirical risk minimization and online learning. In particular, their proof, which employed the classical MacLaurin inequalities, led to improved convergence rate of incremental gradient method (IGM).

Let us recall the famous MacLaurin inequalities for positive real numbers $x_1, \ldots, x_n$ and the normalized $d$-th symmetric sums as
\[ S_d = \binom{n}{d}^{-1} \sum_{\tau \subset \{1, \ldots, n\} \atop |\tau| = d} \prod_{i \in \tau} x_i. \]

where $1 \leq d \leq n$ and $|\tau| :=$ the cardinality of $\tau$. According to the MacLaurin inequalities, we have
\[ S_1 \geq \sqrt[n]{S_2} \geq \sqrt[n]{S_3} \geq \ldots \geq \sqrt[n]{S_n}. \]

+ Partially supported by DMS 1501103 and BigData 1447879.
In particular, $S_1 \geq \sqrt[n]{S_n}$ is the standard AGM inequality. For more details about the classical AGM inequality see [5].

It’s natural to ask whether the MacLaurin inequalities still hold if we replace the real numbers with positive definite matrices. Namely, if $A_1, \cdots, A_n$ are a collection of positive definite operators in some Hilbert space, the MacLaurin mean is defined as

\begin{equation}
\mathcal{P}_d(A) = \frac{(n-d)!}{n!} \sum_{1 \leq j_1, \ldots, j_d \leq n \text{ all different}} A_{j_1} \cdots A_{j_d}
\end{equation}

It is still open whether the norm of $\mathcal{P}_d$ is always less than that of $(\mathcal{P}_1)^n$, i.e.

\begin{equation}
\|\mathcal{P}_d(A)\| \leq C_1 \cdot \|\mathcal{P}_1(A)\|^d
\end{equation}

Ré and Recht show that these noncommutative AGM inequalities hold when there are only two matrices or when all of the matrices commute. Arie, Felix and Rachel [6] prove the inequality holds for products of up to three matrices by a variant of the classic Araki-Lieb-Thirring inequality. Ré and Recht also demonstrate that AGM holds for matrices with a constant depending on the degree $d$ and the dimension $m$ under some assumption. In [1], we prove the (1.2) in both normed and ordered sense under some assumptions.

We define a symmetrized normalized d-th sums as follows:

\begin{equation}
S_d(A) = \frac{(n-d)!}{n!} \sum_{1 \leq j_1, \ldots, j_d \leq n \text{ all different}} A_{j_1}^* A_{j_2}^* \cdots A_{j_d}^* A_{j_d} \cdots A_{j_2} A_{j_1}
\end{equation}

$S_d(A)$ is a second order polynomial. Each term in the summation form of $S_d(A)$ is positive definite automatically. The symmetrized-AGM conjecture may be stated as follows:

\begin{equation}
\|S_d(A)\| \leq C_2 \cdot \|S_1(A)\|^d.
\end{equation}

However, in the symmetrized setting, the following does not hold in general:

\begin{equation}
S_1(A)^d \neq \frac{1}{n^d} \sum_{j_1, \ldots, j_d} A_{j_1}^* A_{j_2}^* \cdots A_{j_d}^* A_{j_d} \cdots A_{j_2} A_{j_1}.
\end{equation}

Therefore Ré and Recht define the with-replacement expectation is defined as

\begin{equation}
\mathbb{E}_{wr,k}[x_{i_1}, \cdots, x_{i_n}] = n^{-k} \sum_{j_1, \ldots, j_k} x_{j_1}^* \cdots x_{j_k}^* x_{j_k} \cdots x_{j_1}.
\end{equation}

That is, we average the value of $f$ over all ordered tuples of elements from $(x_1, \cdots, x_n)$. Similarly, the without-replacement expectation as

\begin{equation}
\mathbb{E}_{wo,k}[x_{i_1}, \cdots, x_{i_n}] = \frac{(n-k)!}{n!} \sum_{j_1 \neq j_2 \neq \cdots \neq j_k} x_{j_1}^* \cdots x_{j_k}^* x_{j_k} \cdots x_{j_1}.
\end{equation}
Ré and Recht also asked for norm or order inequalities of the following:

\begin{align}
\| E_{wo,k}(A) \| & \leq C_3 \cdot \| E_{wr,k}(A) \|, \\
E_{wo,k}(A) & \leq C_4 \cdot E_{wr,k}(A).
\end{align}

The norm symmetric AGM conjecture asks for (1.7) with \( C_3 = 1 \). This still remains open.

**Theorem 1.1.** The order symmetric AGM conjecture (1.8) with \( C_4 = 1 \) is false for \( n \geq 3, d = 3 \).

Therefore, the problem of computing the best constant in (1.4) is particularly interesting. We explore the relation between the sequence \( A \), the length \( d \) and \( C_2 \) as follows:

**Theorem 1.2.** Let \( \{ A_i \} \) be a family of operators on \( B(H) \) satisfying \( \frac{1}{n} \sum_{j=1}^{n} A_j A_j^* = \mathbb{1} \). Then

\begin{align}
(1) & \quad S_1(A) = 1; \\
(2) & \quad (1 - \varepsilon) \cdot S_1(A)^d \leq S_d(A) \leq (1 + \varepsilon) \cdot S_1(A)^d,
\end{align}

where \( \varepsilon = \frac{1 + \sup \| A_k^* A_k \|}{n} \cdot \frac{d(d-1)}{2}, 1 \leq d \leq n \).

Under the condition \( \frac{1}{n} \sum A_j A_j^* = \mathbb{1} \), we obtain

\( S_1(A)^d = E_{wo,d}(A) \) and \( S_d(A) = E_{wr,d}(A) \).

Therefore, we also provide a proof of Ré and Recht’s conjecture with \( C_3 = 1 + \varepsilon \) in (1.7). However for random matrices, the condition \( \frac{1}{n} \sum A_j A_j^* = \mathbb{1} \) is no longer satisfied. Therefore we provide a deviation version of SAGM:

**Theorem 1.3.** Let \( \{ A_i \} \) be a family of i.i.d. random operators such that \( E(A_i^* A_i) = \mathbb{1} \). If \( \left( \int \| \sum A_j^* A_j - E(\sum A_j^* A_j) \|^p_{B(H)} dw \right)^{1/p} \leq n \varepsilon, d \ll n \), then

\begin{align}
(1) & \quad \left( \int \| E_{wo,d}(A) - E(E_{wo,d}(A)) \|^p_{B(H)} dw \right)^{1/p} \approx O(d \varepsilon), \\
(2) & \quad \left( \int \| E_{wo,d}(A) \|^p_{B(H)} dw \right)^{1/p} \leq (1 + O(d \varepsilon)) \left( \int \| E_{wr,d}(A) \|^p_{B(H)} dw \right)^{1/p}
\end{align}

The organization of the paper is as follows: we introduce some basic knowledge about combinatorial theory and von Neumann algebras. In section 2, we try to use some new techniques to estimate the symmetric version of AGM. After that, we give a counter example of the SAGM inequality using free probability. We give the deviation version of SAGM at the end of this section. In section 3, we focus on the Incremental Gradient Method. We also construct three random matrices examples to satisfy the IGM with good constants using group representation theory and spherical design.

2. Upper bounds for symmetrized arithmetic-geometric mean inequalities

We need to recall some definitions from the combinatorial theory of partitions in [2] and [11].
Definition 2.1. Let $\mathbb{P}_n$ be the lattice of all the partitions of $\{1, \ldots, n\}$. For two partitions $\sigma$ and $\pi$, we write $\sigma \leq \pi$ if every block of the partition $\sigma$ is contained in some block of $\pi$ (i.e., any block of the partition $\pi$ can be written as a union of blocks of $\sigma$). In other words, $\pi$ is a refinement of $\sigma$.

Remark 2.2. There are two trivial partitions, $\hat{0}$ and $\hat{1}$, where $\hat{0}$ is the partition into $n$ singletons and $\hat{1}$ is the partition of a single block. Namely, $\hat{0}$ is the smallest partition and $\hat{1}$ is the biggest one in $\mathbb{P}_n$.

Remark 2.3. For a partition $\pi$, let $\nu(\pi)$ be the number of the blocks of the partition $\pi$.

For fixed $d$ we consider the following average symmetric product of noncommutative operators of length $d$: 

$$S_d(A) = \frac{(n-d)!}{n!} \sum_{(i_1, \ldots, i_d) = \hat{0}} A_{i_1}^* A_{i_2}^* \ldots A_{i_d}^* A_{i_d} \ldots A_{i_2} A_{i_1}. $$

Example 2.4. Let $A_i$ be a family of operators such that $\frac{1}{n} \sum A_i^* A_i = 1$.

1. For the partition $[13, 2]$ where $d = 3$ we can have the following upper-bound for the symmetric case of self-adjoint operator.

$$\| \sum_{(i_1, i_2, i_3) = \hat{0}} e_{kj, 1} \otimes A_k A_j A_k \|^2 = \| \sum_{(i_1, i_2, i_3) = \hat{0}} e_{kj, i'j'} \otimes A_{i_1}^* A_{i_2}^* A_k^* A_k A_j A_k \|$$

$$= \| \left( \sum_k e_{kk} \otimes 1 \otimes 1 \otimes A_{i_k}^* \right) \times \left( \sum_j e_{j1} \otimes 1 \otimes A_{i_j}^* \right) \times \left( \sum_k e_{k1} \otimes 1 \otimes 1 \otimes A_{i_k}^* \right) \|$$

$$\times \left( \sum_{k'} e_{k'1} \otimes 1 \otimes 1 \otimes A_{i_{k'}} \right) \times \left( \sum_j e_{j1} \otimes 1 \otimes A_{i_j} \right) \times \left( \sum_{j'} e_{j'1} \otimes 1 \otimes A_{i_{j'}} \right) \|$$

The column or row term is bounded by $\sqrt{n}$, and the diagonal term is bounded by $\sup \|A_{i_{kk}}\|$. Therefore we have

$$\| \sum_{(i_1, i_2, i_3) = \hat{0}} e_{kj, 1} \otimes A_k A_j A_k \|^2 \leq \text{sup} \|A_{i_{kk}}\| (\sqrt{n})^4 \sup \|A_{i_{kk}}\| = n^2 \sup \|A_{i_{kk}}^* A_{i_{kk}}\|.$$ 

2. If we have the partition $[12, 34]$ where $d = 4$ then we have

$$\| \sum_{(i_1, i_2, i_3, i_4) = \hat{0}} e_{kj, 1} \otimes A_k A_j A_k A_j \|^2 \| \sum_{(i_1, i_2, i_3, i_4) = \hat{0}} e_{kj, k'j'} \otimes A_k A_k A_j A_j A_j A_k A_k \|$$

$$= \| \sum_k e_{kk} \otimes 1 \otimes A_{i_k} \| \times \| \sum_k e_{k1} \otimes 1 \otimes A_{i_k} \| \times \| e_{j1} \otimes 1 \otimes A_{i_j} \| \times \| e_{j1} \otimes 1 \otimes A_{i_j} \| \times$$

$$\times \| e_{k'1} \otimes 1 \otimes A_{i_{k'}} \| \times \| e_{k'1} \otimes 1 \otimes A_{i_{k'}} \| \times$$

$$\times \| e_{j'1} \otimes 1 \otimes A_{i_{j'}} \| \times \| e_{j'1} \otimes 1 \otimes A_{i_{j'}} \| \times$$

$$\times \left( \sum_{j'} e_{j'1} \otimes 1 \otimes A_{i_{j'}} \right) \times \left( \sum_{k'} e_{k'1} \otimes 1 \otimes A_{i_{k'}} \right) \times \left( \sum_{k'} e_{k'1} \otimes 1 \otimes A_{i_{k'}} \right) \times \left( \sum_{k'} e_{k'1} \otimes 1 \otimes A_{i_{k'}} \right).$$
Since we have two column terms, two row terms and four diagonal terms, then the symmetric summation has

$$\left\| \sum_{\langle i_1, i_2, i_3, i_4 \rangle = 0} e_{kj,1} \otimes A_k A_k A_j \right\|^2 \leq n^2 \cdot \sup \left\| A_{ik} A_{ik}^* \right\|^2 = n^{\sigma} \cdot \sup \left\| A_{ik} A_{ik}^* \right\|^{d - \nu}.$$

From the above examples we can generalize the upper bound for any partition. By Pisier’s tensor techniques mentioned in [9], we first need to embed each element $a_{ik} \in M$ in larger space $(\otimes B(H)) \otimes M$. For each $i_k \in B_i$ we have

$$Z_{ik} = \begin{cases} 1 \otimes \ldots \otimes e_{kk} \otimes \ldots \otimes 1 \otimes A_{ik} & \text{if } i_k \neq \min B_i \\ 1 \otimes \ldots \otimes e_{1k} \otimes \ldots \otimes 1 \otimes A_{ik} & \text{if } i_k = \min B_i \end{cases}$$

where $\min B_i$ means the smallest index number and $e_{kk}, e_{1k}$ are matrix units in the $i$th component of the tensor form.

Similarly define the adjoint of the operator $Z_{ik}$, denoted as $Z_{ik}^*$, as follow:

$$Z_{ik}^* = \begin{cases} 1 \otimes \ldots \otimes e_{kk} \otimes \ldots \otimes 1 \otimes A_{ik}^* & \text{if } i_k \neq \min B_i \\ 1 \otimes \ldots \otimes e_{1k} \otimes \ldots \otimes 1 \otimes A_{ik}^* & \text{if } i_k = \min B_i \end{cases}$$

**Lemma 2.5.** Let $A_i$ be a family of operators such that $\frac{1}{n} \sum A_i^* A_i = 1$. Denote $\tilde{Z}_{ik} = \sum Z_{ik}$, then we have

$$\left\| \tilde{Z}_{ik} \right\| \leq \begin{cases} \sup m \left\| A_m \right\| & \text{if } i_k \neq \min B_i \\ \frac{n}{\sqrt{n}} & \text{if } i_k = \min B_i \end{cases}$$

and consequently $\tilde{Z}_{ik}^* := \sum Z_{ik}^*$, then we have $\left\| \tilde{Z}_{ik}^* \right\| = \left\| \tilde{Z}_{ik} \right\|$.

**Proof.** If $i_k = \min B_i$, it turns out that $\tilde{Z}_{ik}$ generates a row space. Therefore,

$$\left\| \tilde{Z}_{ik} \right\| = \left\| \tilde{Z}_{ik} \tilde{Z}_{ik}^* \right\|^{1/2} = \left\| (\sum Z_{ik})(\sum Z_{ik}^*) \right\|^{1/2} = \left\| 1 \otimes \ldots \otimes 1 \otimes (\sum A_{ik} A_{ik}^*) \right\|^{1/2} = \left\| \sum A_{ik} A_{ik}^* \right\|^{1/2} = \sqrt{n}$$

If $i_k \neq \min B_i$, it turns out that $\tilde{Z}_{ik}$ generates a matrix form with diagonal component with elements $\{A_{ik}\}$.

$$\left\| \tilde{Z}_{ik} \right\| = \left\| \sum Z_{ik} \right\| = \sup_{i_k} \left\| A_{ik} \right\|.$$

The following is the key lemma for our main result in this section.

**Lemma 2.6.** Let $\sigma$ be a partition in $\mathbb{P}_j$ with the blocks $B_1, \ldots, B_{\nu(\sigma)}$ then we have

$$[\left\| \left[ \sigma \right] \right\| \leq n^{\nu(\sigma)} \cdot \sup \left\| A_{ik}^* A_{ik} \right\|^{\sigma - \nu(\sigma)}.$$
Proof. Let $\sigma = B_1, B_2, \ldots, B_{\nu(\sigma)}$ where $B_i$'s are the blocks in this partition. Then we have

$$
\| \sum_{(i_1, \ldots, i_j) = \sigma} A_{i_1}^* A_{i_2}^* \cdots A_{i_j}^* A_{i_1} \cdots A_{i_{j-1}} A_{i_j} \| 
$$

$$
= \| \sum_{i_1, \ldots, i_j} Z_{i_1}^* Z_{i_2}^* \cdots Z_{i_j}^* Z_{i_1} \cdots Z_{i_{j-1}} A_{i_j} \| 
$$

$$
= \| \sum_{i_j} Z_{i_j}^* \sum_{i_{j-1}} Z_{i_{j-1}}^* \cdots \sum_{i_1} Z_{i_1}^* \sum_{i_{j-1}} Z_{i_{j-1}} \cdots \sum_{i_j} Z_{i_j} \| 
$$

$$
= \| \tilde{Z}_{i_j}^* \cdots \tilde{Z}_{i_2}^* \tilde{Z}_{i_1} \cdots \tilde{Z}_{i_j} \| \leq \prod_{k=1}^j \| \tilde{Z}_{i_k} \| \leq n^\nu(\sigma) \cdot \sup \| A_k^* A_k \|^{j-\nu(\sigma)}.
$$

The last inequality comes from the observation of Lemma 2.5. It explains if we have a block, then there is only one operator in this block with index gives the upper bound norm equal to $n$ and the remaining operators with different indices are bounded by $\sup \| A_k^* A_k \|$. Note that $\nu(\sigma)$ is the number of blocks. In the product form we have $\nu(\sigma)$ items give the bound $n$ and the remaining $j-\nu(\sigma)$ items give the bound $\sup \| A_k^* A_k \|$.  □

Remark 2.7. In particular, we can observe that

1. Let $\sigma$ be 0, $\| [\hat{\sigma}] \| \leq n^k$;
2. Let $\sigma$ be 1, $\| [\hat{\sigma}] \| \leq \sup \| A_k^* A_k \|^k$

Here $k$ means the cardinality of the set.

In this paper without specific explanation, we always assume $\frac{1}{n} \sum A_i^* A_i = 1$, i.e. $S_1(A) = 1$. Then

$$
S_1(A) - S_d(A) = \frac{(n-d)!}{n!} \sum_{(i_1, \ldots, i_d) = \emptyset} (1 - A_{i_1}^* A_{i_2}^* \cdots A_{i_d}^* A_{i_1} \cdots A_{i_2} A_{i_1}).
$$

Most of the techniques we use in our first paper are not working for SAGM. So we need to introduce some other techniques. We need to introduce what Marius’s method ref called the folding technique. Let’s see the following example which explains this technique for $d = 2$.

$$
1 - A_2^* A_1 A_2 = 1 - A_2^* A_2 + A_2^* A_2 - A_2^* A_1^* A_1 A_2 
$$

$$
= (1 - A_2^* A_2) + A_2^*(1 - A_1^* A_1) A_2
$$

Now for $d$ term we have

$$
1 - A_d^* \cdots A_1^* A_1 \cdots A_d 
$$

$$
= (1 - A_d^* A_d) + A_d^*(1 - A_{d-1}^* A_{d-1}) A_d + A_d^* A_d^* (1 - A_{d-2}^* A_{d-2}) A_{d-1} A_d + \ldots
$$

$$
+ A_d^* A_{d-1}^* \cdots A_2^* (1 - A_1^* A_1) A_2 \ldots A_{d-1} A_d
$$

$$
= \sum_{j=1}^d A_d^* \cdots A_{j+1}^* (1 - A_j^* A_j) A_{j+1} \cdots A_d
$$
Given a partition $\sigma$, and $i_1$ is not a singleton in $\sigma$. We denote
\[
[[\sigma]] : = \sum_{(i_1, \ldots, i_j) = \sigma} A^*_{i_j} \cdots A^*_{i_2} (1 - A^*_{i_1} A_{i_1}) A_{i_2} \cdots A_{i_j}.
\]

**Lemma 2.8.** Given a partition $\sigma \in \mathbb{P}_j$ and $i_1$ is not a singleton element in the partition $\sigma$, then
\[
(2.1) \quad \|[\sigma]\| \leq n^{\nu(\sigma)} : C^{|\sigma| - \nu(\sigma)} (1 + \frac{1}{C}),
\]
where $C = \sup \|A^*_k A_k\|$.

**Proof.** Let $\gamma$ be the partition without index $i_1$, then we know $\nu(\sigma) = \nu(\gamma)$, $|\gamma| = j - 1$ and $|\sigma| = j$.

\[
[[\sigma]] = \sum_{(i_1, \ldots, i_j) = \sigma} A^*_{i_j} \cdots A^*_{i_2} (1 - A^*_{i_1} A_{i_1}) A_{i_2} \cdots A_{i_j}
\]
\[
= \sum_{(i_1, \ldots, i_j) = \sigma} A^*_{i_j} \cdots A^*_{i_2} A_{i_2} \cdots A_{i_j} - \sum_{(i_1, \ldots, i_j) = \sigma} A^*_{i_j} \cdots A^*_{i_1} A_{i_1} \cdots A_{i_j}
\]
\[
= \sum_{(i_2, \ldots, i_j) = \gamma} (1 - A^*_{i_1} A_{i_1}) A_{i_2} \cdots A_{i_j} - \sum_{(i_1, \ldots, i_j) = \sigma} A^*_{i_j} \cdots A^*_{i_1} A_{i_1} \cdots A_{i_j}
\]
\[
= [\gamma] - [\sigma]
\]

Then we consider the norm
\[
\|[\sigma]\| \leq \|[\gamma]\| + \|[\sigma]\| \leq n^{\nu(\gamma)} \cdot \sup \|A^*_k A_k\||\gamma| - \nu(\gamma) + n^{\nu(\sigma)} \cdot \sup \|A^*_k A_k\||\sigma| - \nu(\sigma)
\]
\[
\leq n^{\nu(\sigma)} \cdot \sup \|A^*_k A_k\||\sigma| - \nu(\sigma) \left(1 + \frac{1}{\sup \|A^*_k A_k\|}\right).
\]

**Theorem 2.9.** Let $\{A_i\}$ be a family of operator’s on $B(H)$ satisfying $\frac{1}{n} \sum A_i A_i^* = 1$. Then we have
\[
(2.2) \quad \|1 - S_d(A)\| \leq \frac{(1 + C) d(d - 1)}{n \cdot 2}
\]
where $C = \sup \|A^*_k A_k\|$.

**Proof.** Let $i_j = l$ and define $n_l := [1, \ldots, n] - l$ and $n_{l,j-1} := [1, \ldots, n]_{j-1} - l$. Also, let $t$ represents the indices such that if
\[
\text{if } t \notin [n_l] \Rightarrow \text{ some of the indices of } t \text{ are equal to } l.
\]

Then we have
\[
(2.3) \quad 1 - S_d(A) = \sum_{j=1}^d \frac{(n - j)!}{n!} \sum_{(i_1, \ldots, i_j) = \emptyset} A^*_{i_j} \cdots A^*_{i_{j-1}} (1 - A^*_{i_j} A_{i_j}) A_{i_{j-1}} \cdots A_{i_1}
\]
Then let \( l = i_j \)

\[
I_j = \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* (1 - A_l^* A_l) A_{i_{j-1}} \cdots A_{i_1}
\]

\[
= \sum_{i_1} \cdots \sum_{i_{j-1}} \sum_{l=1}^{n} A_{i_1}^* \cdots A_{i_{j-1}}^* (1 - A_l^* A_l) A_{i_{j-1}} \cdots A_{i_1}
\]

\[
= \sum_{l=1}^{n} \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* (1 - A_l^* A_l) A_{i_{j-1}} \cdots A_{i_1}
\]

Then by distributing the sum for both terms and from the given assumption we have

\[
I_j = \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* \sum_{l=1}^{n} (1 - A_l^* A_l) A_{i_{j-1}} \cdots A_{i_1}
\]

\[
= \sum_{l=1}^{n} \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* (1 - A_l^* A_l) A_{i_{j-1}} \cdots A_{i_1}
\]

\[
= - \sum_{l=1}^{n} \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* (1 - A_l^* A_l) A_{i_{j-1}} \cdots A_{i_1}
\]

\[
+ \sum_{l=1}^{n} \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* A_l^* A_i A_{i_{j-1}} \cdots A_{i_1}
\]

Notice that the sum run over the restricted partition \( \{i_1, \ldots, i_{j-1}\} = \hat{0} \) with condition that not all of them in \( n_l^{j-1} \). This will force just one \( i_k = l \), so we have \( (j - 1) \) choices.

\[
\| I - S_d(A) \| \leq \sum_{j=1}^{d} \frac{(n-j)!}{n!} (j-1)! \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* A_{i_{j-1}} \cdots A_{i_1}
\]

\[
+ \sum_{j=1}^{d} \frac{(n-j)!}{n!} (j-1)! \sum_{\{i_1, \ldots, i_{j-1}\} = \hat{0}} A_{i_1}^* \cdots A_{i_{j-1}}^* A_l^* A_i A_{i_{j-1}} \cdots A_{i_1}
\]
Now using the fact that the norm of the restricted partition is less than the full one, i.e. \( \| \langle \sigma \rangle \| \leq \| \| \sigma \|| \| \),
\[
\| \mathbb{1} - S_d(\mathcal{A}) \| \leq \sum_{j=1}^{d} \frac{(n-j)!}{n!} (j-1) \| \sum_{i_k=l} A_{i_1}^* \cdots A_{i_{j-1}}^* A_{i_{j-1}} \cdots A_{i_1} \|
\]
\[
+ C \sum_{j=1}^{d} \frac{(n-j)!}{n!} (j-1) \| \sum_{i_k=l} A_{i_1}^* A_{i_{j-1}}^* A_{i_{j-1}} \cdots A_{i_1} \|
\]
By Remark 2.7 for the partition 0, we obtain
\[
\| \mathbb{1} - S_d(\mathcal{A}) \| \leq \sum_{j=1}^{d} \frac{(n-j)!}{n!} (j-1)n^{j-1} + C \sum_{j=1}^{d} \frac{(n-j)!}{n!} (j-1)n^{j-1}
\]
\[
\leq \sum_{j=1}^{d} \frac{(j-1)}{n} + C \frac{(j-1)}{n} = \frac{(1 + C)}{2} \frac{d(d-1)}{n}
\]
Note that if the condition \( \frac{1}{n} \sum A_i^* A_i = \mathbb{1} \) doesn’t hold, we get
\[
S_1(\mathcal{A})^d = \left( \frac{1}{n} \sum A_i^* A_i \right)^d \neq \frac{1}{n^d} \sum_{j_1, \cdots, j_d} A_{j_1}^* A_{j_2}^* \cdots A_{j_d}^* A_{j_d} \cdots A_{j_1}^* A_{j_1}.
\]
To formulate a sufficient analysis, let us first formalize some notation mentioned in [10]. Throughtout in the remaining of this paper, \([n]\) denotes the set of integers from 1 to \( n \). Let \( \mathbb{D} \) be some domain, \( f : \mathbb{D}^k \to \mathbb{R} \), and \( \mathcal{A} := (x_1, \cdots, x_n) \) a set of \( n \) elements from \( \mathbb{D} \). We define the without-replacement expectation as
\[
E_{wo,k}[f(\mathcal{A})] = \frac{(n-k)!}{n!} \sum_{j_1 \neq (j_2, \cdots, j_k)} f(x_{j_1}, \cdots, x_{j_k}).
\]
That is, we average the value of \( f \) over all ordered tuples of elements from \( (x_1, \cdots, x_n) \). Similarly, the with-replacement expectation is defined as
\[
E_{wr,k}[f(\mathcal{A})] = n^{-k} \sum_{(j_1, \cdots, j_k) = i} f(x_{j_1}, \cdots, x_{j_k}).
\]

**Remark 2.10.** With the above definition (2.4), we take a specific function defined as \( f(x_{j_1}, \cdots, x_{j_k}) := \prod_{j=1}^{k} x_{i_{k-j+1}} \prod_{j=1}^{k} x_{i_j} \). Then \( E_{wo,k}[f(\mathcal{A})] = S_k(\mathcal{A}) \).

Next let’s provide a deviation version of SAGM as follows:

**Theorem 2.11.** Let \( \{A_i\} \) be a family of i.i.d. random operators such that \( E(A_i^* A_i) = \mathbb{1} \). If \( \bigg\| \sum A_i^* A_i - E \left( \sum A_i^* A_i \right) \bigg\| \leq n \varepsilon, d < n \), then
\[
(1) \quad \| E_{wo,d}(\mathcal{A}) - E(E_{wo,d}(\mathcal{A})) \| \approx O(\varepsilon),
\]
\[
(2) \quad \| E_{wo,d}(\mathcal{A}) \| \leq (1 + O(\varepsilon)) \| E_{wo,d}(\mathcal{A}) \|
\]
Proof. (1) Since \( \{A_i\} \) are i.i.d., \( E(\mathbb{E}_{wo,d}) = 1 \). Therefore

\[
\|\mathbb{E}_{wo,d}(A) - E(\mathbb{E}_{wo,d})\| = \|1 - \mathbb{E}_{wo,d}(A)\| \leq 
\]

(2.5)

\[
\leq \sum_{j=1}^{d} \frac{(n-j)!}{n!} \sum_{l \in [n]} \sum_{(i_1, \ldots, i_{j-1}) = \emptyset, i_j = l, k \in [n], k \neq j} A_{i_1}^* \ldots A_{i_{j-1}}^* (1 - A_{i_j}^* A_{i_j}) A_{i_{j-1}} \ldots A_{i_1}
\]

(2.6)

\[
+ \sum_{j=1}^{d} \frac{(n-j)!}{n!} \sum_{l \in [n]} \sum_{(i_1, \ldots, i_{j-1}) = \emptyset, i_j = l, k \in [n], k \neq j} A_{i_1}^* \ldots A_{i_{j-1}}^* A_{i_{j-1}} \ldots A_{i_1}
\]

(2.7)

\[
+ \sum_{j=1}^{d} \frac{(n-j)!}{n!} \sum_{l \in [n]} \sum_{(i_1, \ldots, i_{j-1}) = \emptyset, i_j = l, k \in [n], k \neq j} A_{i_1}^* \ldots A_{i_{j-1}}^* A_{i_{j-1}} \ldots A_{i_1}
\]

The term on the left side of the inequality (2.4) can factor out \( \sum (1 - A_{i_j}^* A_{i_j}) \) since the index \( i_j \) is independent with the others. And the norm in (2.5) is invariant with respect to the index \( l \). (2.6) is different with (2.5) with a constant \( C \), after factoring out the \( \|A_{i_j}^* A_{i_j}\| \).

\[
\|\mathbb{E}_{wo,d}(A) - E(\mathbb{E}_{wo,d}(A))\|
\]

\[
\leq \sum_{j=1}^{d} \frac{(n-j)!}{n!} \sum_{[i_1, \ldots, i_{j-1}] = \emptyset} A_{i_1}^* \ldots A_{i_{j-1}}^* A_{i_{j-1}} \ldots A_{i_1} \cdot \|\sum_{i_j} (1 - A_{i_j}^* A_{i_j})\|
\]

\[
+ (1 + C) \sum_{j=1}^{d} \frac{(n-j)!}{n!} (j-1) \sum_{[i_1, \ldots, i_{j-1}] = \emptyset} A_{i_1}^* \ldots A_{i_{j-1}}^* A_{i_{j-1}} \ldots A_{i_1}
\]

Therefore we get

\[
\|\mathbb{E}_{wo,d}(A) - E(\mathbb{E}_{wo,d}(A))\| \leq \sum_{j=1}^{d} \frac{(n-j)!}{n!} \frac{n!}{(n-j+1)!} \|\mathbb{E}_{wo,d}^{-1}(A)\| \cdot \left( n \varepsilon + (1 + C) \cdot (j-1) \right)
\]
Then by using \( \| \mathbb{E}_{w_0,j-1}(A) - \mathbb{E}_{w_0,j-1}(A) - E(\mathbb{E}_{w_0,j-1}(A)) + E(\mathbb{E}_{w_0,j-1}(A)) \| \), we have

\[
\| \mathbb{E}_{w_0,d}(A) - E(\mathbb{E}_{w_0,d}(A)) \| \leq \sum_{j=1}^{d} \frac{n\varepsilon + (1 + C) \cdot (j - 1)}{n - j + 1} \| \mathbb{E}_{w_0,j-1}(A) - E(\mathbb{E}_{w_0,j-1}(A)) \|
\]

Using the discrete case of Grönwall’s lemma [4],

\[
\| \mathbb{E}_{w_0,d}(A) - E(\mathbb{E}_{w_0,d}(A)) \| \leq f(d) \exp(f(d)), f(d) := \sum_{j=1}^{d} \frac{n\varepsilon + (1 + C) \cdot (j - 1)}{n - j + 1}
\]

It remains to find an upper bound for \( f(d) \)

\[
f(d) = \sum_{j=1}^{d} \frac{(n - j + 1)\varepsilon + (\varepsilon + 1 + C) \cdot (j - 1)}{n - j + 1} = \sum_{j=1}^{d} (\varepsilon + (\varepsilon + 1 + C) \cdot \frac{j - 1}{n - j + 1})
\]

\[
\leq d\varepsilon + (\varepsilon + 1 + C) \sum_{j=1}^{d} \frac{j - 1}{n - j + 1} \leq d\varepsilon + (\varepsilon + 1 + C) \cdot \frac{d(d - 1)}{n - d + 1}
\]

Since \( d \ll n, f(d) \approx O(d\varepsilon) \), then \( \| \mathbb{E}_{w_0,d}(A) - E(\mathbb{E}_{w_0,d}(A)) \| \approx O(d\varepsilon) \).

(2) Since \( \{A_i\} \) are i.i.d., \( E(\mathbb{E}_{w_0,d}) = 1 \).

\[
\mathbb{E}_{w_0,d} - E(\mathbb{E}_{w_0,d}) = \frac{1}{n^d} \sum_{j_1,\ldots,j_d} x_{j_1}^* \cdots x_{j_d}^* x_{j_d} \cdots x_{j_1} - 1
\]

(2.8)

\[
= \frac{1}{n^{d-1}} \sum_{j_1,\ldots,j_{d-1}} x_{j_1}^* \cdots x_{j_{d-1}}^* \left( \frac{1}{n^d} \sum_{j_d} x_{j_d} x_{j_d} - 1 \right) x_{j_{d-1}} \cdots x_{j_1} + \frac{1}{n^{d-1}} \sum_{j_1,\ldots,j_{d-1}} x_{j_1}^* \cdots x_{j_{d-1}}^* x_{j_{d-1}} \cdots x_{j_1} - 1
\]

Thanks to \( \| \sum A_i^* A_i - E \left( \sum A_i^* A_i \right) \| \leq n\varepsilon \) and Hölder inequality, the term (2.8) admits the following estimate

\[
\| \sum_{j_1,\ldots,j_{d-1}} x_{j_1}^* \cdots x_{j_{d-1}}^* \left( \frac{1}{n^d} \sum_{j_d} x_{j_d} x_{j_d} - 1 \right) x_{j_{d-1}} \cdots x_{j_1} \| \leq \varepsilon \| \sum_{i_1=1}^{n} x_{i_1}^* x_{i_1} \|^{d-1}
\]

Since \( \| \sum_{i_1=1}^{n} x_{i_1}^* x_{i_1} \| \leq n(1 + \varepsilon) \) and triangle inequality, we get

\[
\| \mathbb{E}_{w_0,d} - E(\mathbb{E}_{w_0,d}) \| \leq \frac{1}{n^{d-1}} \varepsilon \cdot (n(1 + \varepsilon))^{d-1} + \| \mathbb{E}_{w_0,d-1} - E(\mathbb{E}_{w_0,d-1}) \|
\]

Denote \( a_d = \| \mathbb{E}_{w_0,d} - E(\mathbb{E}_{w_0,d}) \| \), we have a sequence \( \{a_d\} \) has \( a_1 \leq \varepsilon \) and the iteration inequality,

\[
a_d \leq \varepsilon(1 + \varepsilon)^{d-1} + a_d
\]
Then such that Lemma 3.3.

First, we need to prove two main lemmas. Then by the triangle inequality, we obtain

\[ \|E_{w_0,d}(A) - E(E_{w_0,d}(A))\| \leq \sum_{i=1}^{d} \varepsilon(1 + \varepsilon)^{i-1} \leq O(d\varepsilon) \]

Then by triangle inequality, we obtain

\[ \|E_{w_0,d}(A) - E(E_{w_0,d}(A))\| = \|E_{w_0,d}(A) - E(E_{w_0,d}(A)) - (E_{w_0,d}(A) - E(E_{w_0,d}(A)))\| \]

\[ \leq \|E_{w_0,d}(A) - E(E_{w_0,d}(A))\| \|E_{w_0,d}(A) - E(E_{w_0,d}(A))\| \approx O(d\varepsilon) \]

3. A Counterexample for the Symmetric Arithmetic Geometric Mean Inequality

In this section, we provide an example from free probability [12] which proves that symmetric arithmetic geometric mean inequality (SAGM) is not true in general.

Let’s recall the construction of the reduced amalgamated free product of von Neumann algebras. Let \(A_1, \ldots, A_n\) be a family of von Neumann algebras and let \(\langle a \rangle\) be a common von Neumann subalgebra of \(A_k\) generated by an element \(a\). We will assume that there is a normal faithful conditional expectation \(E_k : A_k \rightarrow \langle a \rangle\) for each \(k\). Let \(A = \ast_{(a)} A_k\) be the reduced amalgamated free product of \(A_1, \ldots, A_n\) over \(\langle a \rangle\) with respect to the \(E_k\). We are concerned about the case when \(d = 3\).

**Fact 3.1.** We list some properties for the von Neumann algebras \(A_j := \langle u_j, a \rangle\) generated by \(u_j\) and \(a\):

1. \(A_j\) is freely independent over \(A\), then \(E_{A_j}(X) = E_A(X)\).
2. If \(X = b_1b_2\ldots b_n\) with \(b_i \in A_j, b_i \in A\), then \(E_A(X) = 0\).
3. \(E_A(u_kab) = \tau(u_k)ab = 0\) for all \(a, b \in A\).
4. \(\tau(c_{j_1}\ldots c_{j_n}) = 0, c_{j_i} \in \hat{A_j}, j_i \in \{1, \ldots, n-1\}, c_{j_n} \in A_n\) and \(j_1 \neq j_2 \neq \cdots \neq j_n\).

**Theorem 3.2.** Let \(u_j\) be unitaries and \(\{u_1, \ldots, u_n, a\}\) be freely independent operators such that

\[ a_j = au_j, \ \tau(u_j) = \tau(a) = 0, \ \tau(a^2) = 1 \text{ and } a^2 \neq 1. \]

Then

\[ E_{w_0,3}(a_{j_1}, a_{j_2}, a_{j_3}) = \frac{1}{n(n-1)(n-2)} \sum_{\langle j_1, j_2, j_3 \rangle = 0} a_{j_1}a_{j_2}a_{j_3}a_{j_2}^*a_{j_3}^*a_{j_1}^* \]

\[ E_{w_3,3}(a_{j_1}, a_{j_2}, a_{j_3}) = \frac{1}{n^3} \sum_{\langle j_1, j_2, j_3 \rangle = 1} a_{j_1}a_{j_2}a_{j_3}a_{j_2}^*a_{j_3}^*a_{j_1}^* \]

do not satisfy

\[ E_{w_0,3}(a_{j_1}, a_{j_2}, a_{j_3}) \leq E_{w_3,3}(a_{j_1}, a_{j_2}, a_{j_3}). \]

In this following, we will replace \(E_{w_0,3}\) with \(E_{wo}\) (respectively replace \(E_{w_3,3}\) with \(E_{wr}\)).

First, we need to prove two main lemmas.

**Lemma 3.3.** \(\tau(E_{wo}(a_{j_1}, a_{j_2}, a_{j_3})) = \tau(E_{wr}(a_{j_1}, a_{j_2}, a_{j_3})) = \tau(a^2)^3\)
Proof. We will use the free independent condition which says that \( \tau(a_{i_1}a_{i_2}...a_{i_n}) = 0 \) if \( i_1 \neq i_2 \neq \cdots \neq i_n \) and \( a = a - \tau(a) \). We claim that for all choices \( j_1, j_2, j_3 \), we have
\[
(3.1) \quad \tau(a_{j_1}a_{j_2}a_{j_3}a_{j_1}^*a_{j_2}^*a_{j_1}^*) = \tau(a^2)^3
\]
Let us start by using the definition of the operator \( a_{ij} \) frequently.
\[
\tau(a_{j_1}a_{j_2}a_{j_3}a_{j_1}^*a_{j_2}^*a_{j_1}^*) = \tau(a_{j_1}a_{j_2}a_{j_2}^*a_{j_1}^*)
\]
\[
= \tau(a_{j_1}a_{j_2}(\hat{a}^2)a_{j_2}^*a_{j_1}^*) + \tau(a_{j_1}a_{j_2}a_{j_2}^*a_{j_1}^*)\tau(a^2)
\]
\[
= \tau(a_{j_1}a_{j_2}(\hat{a}^2)a_{j_2}^*a_{j_1}^*) + \tau(a_{j_1}a_{j_2}a_{j_2}^*a_{j_1}^*)\tau(a^2)
\]
\[
= \tau(a_{j_1}a_{j_2}(\hat{a}^2)a_{j_2}^*a_{j_1}^*) + \tau(\hat{a}^2a_{j_2}a_{j_1}^*)\tau(a^2) + \tau(a^2)\tau(a_{j_1}^*)\tau(a^2)
\]
\[
= \tau(a_{j_1}a_{j_2}(\hat{a}^2)a_{j_2}^*a_{j_1}^*) + \tau(\hat{a}^2u_{j_2}^*u_{j_1}^*)\tau(a^2) + \tau(\hat{a}^2u_{j_2}u_{j_1})\tau(a^2) + \tau(a^2)\tau(a_{j_1}^*)\tau(a^2)
\]
\[
= \tau(a_{j_2}(\hat{a}^2)a_{j_2}^*a_{j_1}^*) + \tau(\hat{a}^2)\tau(a_{j_1}^*)\tau(a^2)
\]
\[
= \tau(a_{j_2}(\hat{a}^2)a_{j_2}^*a_{j_1}^*) + \tau(\hat{a}^2)\tau(a_{j_1}^*)\tau(a^2)
\]
\[
= \tau(\hat{a}^2a_{j_2}u_{j_1}^2u_{j_1}) + \tau(u_{j_2}a(\hat{a}^2)u_{j_2}^*a)\tau(a^2) + \tau(\hat{a}^2)\tau(a_{j_1}^*)\tau(a^2)
\]
\[
= \tau(\hat{a}^2)\tau(a_{j_1}^*)\tau(a^2) = \tau(a^2)^3 = 1. \qedhere
\]

The above zero terms hold thanks to freeness. We will use the following folding techniques introduced by Junge in section 6 of [7] to write \( 1 - \mathbb{E}_{wo} \) and \( 1 - \mathbb{E}_{wr} \) for the operators \( a_j \) where \( a_j = au_j \). Recall that
\[
1 - a_{j_1}a_{j_2}a_{j_3}a_{j_2}^*a_{j_1}^* = 1 - a_{j_1}a_{j_1}^* + a_{j_1}(1 - a_{j_2}a_{j_3}a_{j_3}^*)a_{j_1}^*
\]
\[
= 1 - a_{j_1}a_{j_1}^* + a_{j_1}(1 - a_{j_2}a_{j_2}^*)a_{j_1}^* + a_{j_1}a_{j_2}(1 - a_{j_3}a_{j_3}^*)a_{j_2}a_{j_1}^*.
\]
Then we have
\[
1 - \mathbb{E}_{wo}(a_1, a_j, a_k) = \frac{1}{n} \sum_{j=1}^{n} a_j^2 + \frac{1}{n(n-1)} \sum_{j=1, j\neq k}^{n} a_j a_k a_k^* a_j
\]
\[
+ \frac{1}{n(n-1)(n-2)} \sum_{j\neq k\neq l} a_j a_k (1 - a_i a_i^*) a_k a_j
\]
\[
= (1 - a_j^2) + \frac{1}{n} \sum_{j=1}^{n} a_j (1 - a_j^2) a_j^* + \frac{1}{n(n-1)} \sum_{j\neq k} a_j a_k (1 - a_j^2) a_k^* a_j
\]
\[
= (1 - a_j^2) + \frac{1}{n} \sum_{j} a_j (1 - a_j^2) a_j^* + \frac{1}{n(n-1)} \sum_{j, k} a_j a_k (1 - a_j^2) a_k^* a_j
\]
\[
- \frac{1}{n(n-1)} \sum_{j} a_j^2 (1 - a_j^2) (a_j^*)^2
\]
(3.2)
and for the expectation for with-replacement

\[(3.3) \quad 1 - \mathbb{E}_{wr}(a_l, a_j, a_k) = (1 - a^2) + \frac{1}{n} \sum_{j=1}^{n} a_j (1 - a^2) a^*_j + \frac{1}{n^2} \sum_{j \neq k} a_j a_k (1 - a^2) a^*_k a^*_j\]

Then we take the difference between \(1 - \mathbb{E}_{wo}\) and \(1 - \mathbb{E}_{wr}\), then we have

\[\mathbb{E}_{wo} - \mathbb{E}_{wr} = 1 - \mathbb{E}_{wr} - (1 - \mathbb{E}_{wo})\]

\[= \left[ \frac{1}{n^2} - \frac{1}{n(n-1)} \right] \sum_{j,k} a_j a_k (1 - a^2) a^*_k a^*_j + \frac{1}{n(n-1)} \sum_{j=1}^{n} a_j (a_j (1 - a^2) a^*_j) a^*_j\]

The next Lemma shows that some terms will vanish under the conditional expectation \(E_{A_j}\).

**Lemma 3.4.** The conditional expectation \(E_{A_j}\) has the following properties:

(i) \(E_{A_j}(a_k^2(1 - a^2)(a^*_k)^2) = 0 \) for \(k \neq j\)

(ii) \(E_{A_j}(a_k a_l(1 - a^2)a^*_l a^*_k) = 0 \) unless \(k = l = j\)

**Proof.** Let us start by (i) and assume \(k \neq j\),

\[E_{A_j}(a_k a_l(1 - a^2) a^*_k a^*_l a) = a E_{A_j}(u_k a u_k (1 - a^2) u^*_k a^*_k) a\]

\[= a E_{A}(u_k a u_k (1 - a^2) u^*_k a^*_k) a = 0\]

For (ii) we will consider the following cases for \(j, l, k:\)

1. If \(j \neq k \neq l\) then \(E_{A_j}(a_k a_l(1 - a^2)a^*_l a^*_k) = 0\)
2. If \(j = k\) then \(E_{A_j}(a_k a_l(1 - a^2)a^*_l a^*_k) = 0\)
3. If \(j = l\) and \(k \neq l\) \(E_{A_j}(a_k a_l(1 - a^2)a^*_l a^*_k) = 0\)

In case (1), we deduce from above

\[E_{A_j}(a_k a_l(1 - a^2)a^*_l a^*_k) = E_{A_j}(u_k a u_l (1 - a^2) u^*_l a^*_k) = a E_{A_j}(u_k a u_l (1 - a^2) u^*_l a^*_k) a\]

Indeed let \(\omega \in A_j\) and \(\omega := E_{A}(\omega) + \omega - E_{A}(\omega) = E_{A}(\omega) + \hat{\omega}\), then

\[E_{A_j}(u_k a u_l (1 - a^2) u^*_l a^*_k) = \tau(u_k a u_l (1 - a^2) u^*_l a^*_k \omega) = \tau(u_k a u_l (1 - a^2) u^*_l a^*_k \omega) = \tau(u_k a u_l (1 - a^2) u^*_l a^*_k \omega) = \tau(u_k a u_l (1 - a^2) u^*_l a^*_k \omega) = 0\]

by (2) and (4) in Remark 3.1

In case (2) we have

\[(3.4) \quad E_{A_j}(a u_j a u_l (1 - a^2) u^*_l a^*_j a) = a u_j a E_{A_j}(u_l (1 - a^2) u^*_l) a u^*_j a = 0.\]

In case (3), when \(j = l\) and \(k \neq l\), we have

\[(3.5) \quad E_{A_j}(a u_k a u_j (1 - a^2) u^*_k a^*_j a) = a E_{A_j}(u_k a u_j (1 - a^2) u^*_j a^*_k) a = 0\]

Since it’s the same with case (1).

Now we can prove the theorem by using the two lemma above.
Proof of Theorem 3.2. Assume $n \geq 2$ and $E_{wo}(A) \leq E_{wr}(A)$. Then by Lemma 3.3
\[ \tau(E_{wo}(A)) = \tau(E_{wr}(A)) = \tau(a^2)^3 = 1 \]
implies
\[ \tau(E_{wr}(A) - E_{wo}(A)) = 0 \]
Since $\tau$ is faithful, then
\[ x \geq 0, \tau(x) = 0 \implies x = 0. \]
We get
\[ E_{wo}(A) = E_{wr}(A). \]
Therefore,
\[ E_{A_j}(E_{wo} - E_{wr}) = 0. \]
Indeed
\[
E_{A_j}(E_{wo} - E_{wr}) = \frac{-1}{n^2(n-1)}a_j^2(1-a^2)(a_j^*)^2 + \frac{1}{n(n-1)}a_j^2(1-a^2)(a_j^*)^2
\]
\[ = \frac{1}{n^2}a_j^2(1-a^2)(a_j^*)^2. \]
Hence
\[ E_{A_j}(a_j^2(1-a^2)(a_j^*)^2) = 0 \]
which means
\[ a_j^2(1-a^2)(a_j^*)^2 = 0. \]
Therefore
\[ a^2 = 1. \]
Then we get contradiction with the assumption. \[ \blacksquare \]
Following Theorem 2.11 we may replace the $\|\cdot\|$ with operator norm for free products, we have the following:

**Corollary 3.5.** Let $A = \{a_i\}_{i=1}^n$ be freely independent operators such that $\tau(a_i^*a_i) = 1$. If \( \frac{1}{n} \| \sum a_i^*a_i - \tau(\sum a_i^*a_i) \| \leq \varepsilon \) and $d \ll n$, then
\[ (1) \| E_{wo,d}(A) - E(E_{wo,d}(A)) \| \leq O(d\varepsilon), \]
\[ (2) \| E_{wo,d}(A) \| \leq (1 + O(d\varepsilon)) \| E_{wr,d}(A) \| \]

**Remark 3.6.** In comparison with Theorem 3.2, a family of freely independent operator allows some tolerance between $\sum a_i^*a_i$ and $\tau(\sum a_i^*a_i)$, then the symmetrized AGM still holds with a constant $1 + O(d\varepsilon)$. Moreover, when $n$ is closer and closer to infinity, $1 + O(d\varepsilon)$ gets close to 1. Then we obtain the symmetrized-AGM inequality.

4. **Incremental Gradient Method and the symmetric AGM inequality**

In this section we consider the Incremental Gradient Method (IGM) in higher dimensions where we demonstrate the error for sampling without replacement-method by using the upper bound of the symmetric AGM inequality.
4.1. Recursion formula for IGM. Let \( x_s \) be a vector in \( \mathbb{R}^m \) and set
\[
y_i = a_i^* x_s + w_i, \text{ for } i = 1, \ldots, n
\]
where \( a_i \in \mathbb{R}^m \) are test vectors and \( w_i \) are i.i.d. Gaussian random variables with mean zero and variance \( \rho^2 \).

We compare sampling with-replacement versus without-replacement sampling for IGM on the cost function
\[
(4.1) \quad \min_x \sum_{i=1}^{n} (a_i^* x - y_i)^2.
\]
Suppose we walk over \( k \) steps of IGM with constant step size \( \gamma \) and we access the terms \( i_1, \ldots, i_k \) in that order. Then we get
\[
x_{i_k} = x_{i_{k-1}} - \gamma a_{i_k} (a_{i_k}^* x_{i_{k-1}} - y_{i_k}) = (I - \gamma a_{i_k} a_{i_k}^*) x_{i_{k-1}} + \gamma a_{i_k} y_{i_k}
\]
Subtracting \( x_s \) from both sides of this equation, then gives
\[
x_{i_k} - x_s = (I - \gamma a_{i_k} a_{i_k}^*) x_{i_{k-1}} + \gamma a_{i_k} y_{i_k} - x_s = (I - \gamma a_{i_k} a_{i_k}^*) x_{i_{k-1}} + \gamma a_{i_k} (a_{i_k}^* x_s + w_{i_k}) - x_s = (I - \gamma a_{i_k} a_{i_k}^*)(x_{i_{k-1}} - x_s) + \gamma a_{i_k} w_{i_k}
\]
Substitute by
\[
x_{i_{k-1}} = (I - \gamma a_{i_{k-1}} a_{i_{k-1}}^*) x_{i_{k-2}} + \gamma a_{i_{k-1}} y_{i_{k-1}} \text{ and } y_{i_{k-1}} = a_{i_{k-1}}^* x_s + w_{i_{k-1}},
\]
then we have
\[
x_{i_k} - x_s = (I - \gamma a_{i_k} a_{i_k}^*)(I - \gamma a_{i_{k-1}} a_{i_{k-1}}^*) x_{i_{k-2}} + \gamma a_{i_{k-1}} (a_{i_{k-1}}^* x_s + w_{i_{k-1}}) - x_s + \gamma a_{i_k} w_{i_k} = (I - \gamma a_{i_k} a_{i_k}^*)(I - \gamma a_{i_{k-1}} a_{i_{k-1}}^*) x_{i_{k-2}} - (I - \gamma a_{i_{k-1}} a_{i_{k-1}}^*) x_s + \gamma a_{i_{k-1}} w_{i_{k-1}} + \gamma a_{i_k} w_{i_k} = (I - \gamma a_{i_k} a_{i_k}^*)(I - \gamma a_{i_{k-1}} a_{i_{k-1}}^*) x_{i_{k-2}} - x_s + (I - \gamma a_{i_k} a_{i_k}^*) \gamma a_{i_{k-1}} w_{i_{k-1}} + \gamma a_{i_k} w_{i_k}.
\]
By iteration over \( k \), we obtain the following term
\[
(4.2) \quad x_{i_k} - x_s = \prod_{j=1}^{k} (I - \gamma a_j a_j^*) (x_0 - x_s) + \sum_{l=1}^{k} \left[ \prod_{k \geq j \geq l} (I - \gamma a_j a_j^*) \gamma a_l \right] w_l
\]
4.2. From incremental gradient method to symmetric-AGM. Now we take the inner product \( \langle x_k - x_s, x_k - x_s \rangle = \|x_k - x_s\|^2 \) and take a partial expectation with respect to \( w_i \) to estimate \( \|x_k - x_s\| \) after \( k \) steps mentioned in (4.2). If \( x, y \) are mutually independent random vectors such that \( \mathbb{E}(y) = 0 \), then we have \( \mathbb{E}(x + y, x + y) = \mathbb{E}((x, x)) + \mathbb{E}(y, y) \).
Because of this property, taking the expectation of (4.2), we get

\[ E_{w_0}[\|x_k - x_s\|^2] = E_{w_0} \left[ \| \prod_{j=1}^{k} (I - \gamma a_{ij} a_{ij}^*) (x_0 - x_s) \|^2 \right] + \rho^2 \gamma^2 \sum_{l=1}^{k} E_{w_0} \left[ \| \prod_{k \geq j > l} (I - \gamma a_{ij} a_{ij}^*) a_{ij} \|^2 \right] \]

Denote \( A_{ij} := I - \gamma a_{ij} a_{ij}^* \), then \( \{A_{ij}\} \) be a family of a self adjoint operators and the above equation can be written as

\[ E_{w_0}[\|x_k - x_s\|^2] = E_{w_0} \left[ \| \prod_{j=1}^{k} A_{ij} (x_0 - x_s) \|^2 \right] + \rho^2 \gamma^2 \sum_{l=1}^{k} E_{w_0} \left[ \| \prod_{k \geq j > l} A_{ij} a_{ij} \|^2 \right]. \]

Expanding the square norm in the \( E_{w_0} \), we get

\[ E_{w_0}[\|x_k - x_s\|^2] \leq E_{w_0} \left[ \| A_k^* \cdots A_1^* A_1 \cdots A_k \| \right] \|x_0 - x_s\|^2 + \rho^2 \gamma^2 \sum_{l=1}^{k} E_{w_0} \left[ \| A_k^* \cdots A_{l+1}^* A_{l+1} \cdots A_k \| \right] \|a_i\|^2 \]

Here we use \( S_d(A) \) to replace the symbol \( E_{w_0} \left[ \| A_k^* \cdots A_1^* A_1 \cdots A_k \| \right] \). We have to split the two terms of (4.4) into \( l \leq k - 1 \) and \( l = k \):

\[ E_{w_0}[\|x_k - x_s\|^2] \leq \| S_d(A) \| \cdot \|x_0 - x_s\|^2 + \rho^2 \gamma^2 \left( \sum_{l=1}^{k-1} E_{w_0} \left[ \| A_k^* \cdots A_{l+1}^* A_{l+1} \cdots A_k \| \right] + 1 \right) \sup_{j \in [n]} \|a_j\|^2 \]

Note that the term \( I_l \) misses out some indexes and hence is not exactly \( S_l \). We define \( S_{j,k}(A) = E_{w_0} \left[ A_k^* \cdots A_{j+1}^* A_{j+1} \cdots A_k \right] \) where \( j < k \). Then (4.5) can be reformulated as follows:

\[ E_{w_0}[\|x_k - x_s\|^2] \leq \| S_k(A) \| \cdot \|x_0 - x_s\|^2 + \rho^2 \gamma^2 \left( \sum_{l=1}^{k-1} \| S_{j,k}(A) \| + 1 \right) \sup_{j \in [n]} \|a_j\|^2 \]

Observe that

1. \( S_{l,k} = \text{avg} \left( \sum_{i_1, \ldots, i_{k-l} \in [n]-[l]} A_{i_{k-l}}^* \cdots A_{i_1}^* A_{i_1} \cdots A_{i_{k-l}} \right) \), and

2. \( S_{l,k} \leq \sum A_{i_{k-l}}^* S_{l+1,k} A_{i_{k-l}} \).

Lemma 4.1.

\[ S_{l,k} \leq C_{k,l} S_{k-l}, \quad C_{k,l} := \frac{(n \ldots n - l + 1)(n \ldots n - (k - l) + 1)}{n \ldots n - k + 1}. \]
Proof. For fixed \(l\), \(E_{w_0} A^*_k \ldots A^*_{i_{l+1}} A_{i_{l+1}} \ldots A_{i_k}\) is comparable to \(S_{k-l}\). Thanks to the above observation,

\[
S_{l,k} = \sum_{i_{k},...,i_{l+1} \in \{1,...,n\} - \{i_1,...,i_l\}} A^*_k \ldots A^*_{i_{l+1}} A_{i_{l+1}} \ldots A_{i_k} 
\]

\[
\leq \sum_{i_{k},...,i_{l+1} \in \{1,...,n\}} A^*_k \ldots A^*_{i_{l+1}} A_{i_{l+1}} \ldots A_{i_k} 
\]

\[
\leq \frac{1}{n...(n-l+1)} \sum_{i_{1},...,i_{l}} n...(n-(k-l)+1)S_{k-l} 
\]

\[
= \left(\frac{n..n-l+1}{n..n-k+1}\right) \frac{n^l n^{k-l}}{n^k} S_{k-l}. \tag*{\blacksquare}
\]

**Lemma 4.2.** The constant \(C_{k,l}\) has the following property:

1. \(C_{k,l} \simeq \exp\left(\frac{lk}{n-k}\right)\).
2. \(C_{k,k-l} = C_{k,i}\)

Proof. The (2) is trivial based on (1). So we just need to verify the part (1). Multiply \(\frac{1}{n^k}\) to the fraction form, we get

\[
C_{k,l} = \frac{n...(n-l+1)}{n...(n-k+1)} \frac{n^{n-l}}{n^{n-k}} \cdot \frac{n^l n^{k-l}}{n^k} \to 1
\]

Let us discuss one of these terms

\[
\alpha_k = \ln \frac{n..n-k+1}{n^k} = \sum_{j=0}^{k-1} \ln \frac{n-j}{n} = \sum_{j=0}^{k-1} \ln(1 - \frac{j}{n}).
\]

Taking the logarithm of this constant \(C_{k,l}\), we will have \(\alpha_l + \alpha_{k-l} - \alpha_k\) such that

\[
\alpha_l + \alpha_{k-l} - \alpha_k = \sum_{j=0}^{l-1} \ln(1 - \frac{j}{n}) + \sum_{j=0}^{k-1} \ln(1 - \frac{j}{n}) - \sum_{j=0}^{k-1} \ln(1 - \frac{j}{n})
\]

\[
= \sum_{j=0}^{k-l-1} \ln(1 - \frac{j}{n}) - \sum_{j=l}^{k-1} \ln(1 - \frac{j}{n}) = \sum_{j=0}^{k-l-1} \ln \left(1 + \frac{l}{n-l-j}\right)
\]

\[
\leq (k-l) \ln(1 + \frac{l}{n-k}) = \ln(1 + \frac{l}{n-k})^{k-l}
\]

Therefore

\[
C_{k,l} = \exp(\alpha_l + \alpha_{k-l} - \alpha_k) \leq (1 + \frac{l}{n-k})^{k-l} \leq \exp\left(\frac{l(k-l)}{n-k}\right) \tag*{\blacksquare}
\]

4.3. Convergent rate for the IGM.

**Theorem 4.3.** For the IGM, \(w_i\) are i.i.d. Gaussian random variables with mean zero and variance \(\rho^2\), and the test vectors \(a_i\) satisfy
(1) $\frac{1}{n} \sum_{j=1}^{n} a_{ij} a_{ij}^* = \sigma I$,
(2) $\sup \|a_{ij}\|^2 = \mu$.

For $k$ iterations, we will have the following estimate

$$E_{\omega_0}[^{\|x_k - x_\star\|^2}] \leq \varphi^k \left( 1 + \frac{k(k - 1)}{2n} (1 + C_1) \right) \eta$$

Here $\phi = 1 - 2\gamma\sigma + \gamma^2\sigma\mu$, $\beta = \|x_0 - x_\star\|$, $C_1, C_2$ are two constants related to $\gamma, \mu, \sigma$.

From now on, we always assume $\frac{1}{n} \sum_{j=1}^{n} a_{ij} a_{ij}^* = \sigma I$, $\sup \|a_{ij}\|^2 = \mu$. Recall that we define $A_j$ as $I - \gamma a_{ij} a_{ij}^*$, and obtain the following relations

(1) $A_j^* A_j = I - 2\gamma a_{ij} a_{ij}^* + \gamma^2 a_{ij} a_{ij}^* \|a_{ij}\|^2$
(2) $\frac{1}{n} \sum_{j=1}^{n} A_j^* A_j = (1 - 2\gamma\sigma + \gamma^2\sigma\mu) \cdot I := \varphi \cdot I$
(3) $\sup \|A_j\|^2 \leq \max \{|1 - \gamma\mu|, 1\}$

By rescaling the operators $A_j = \sqrt{\varphi} U_j$ in the average form (2) and using the upper bound inequality (2.2) for $\frac{1}{n} \sum U_j^2 U_j = I$, we can get

$$\|S_k(A)\| \leq \varphi^k \|S_k(U)\| \leq \varphi^k \left( 1 + \frac{k(k - 1)}{2n} (1 + C) \right) = \varphi^k \left( 1 + \Delta_k \right)$$

Here

$$C = \sup \|A_j\|^2, \quad \Delta_k = \frac{k(k - 1)}{2n} (1 + C).$$

Now, thanks to the inequality (4.6), we have

$$E_{\omega_0}[^{\|x_k - x_\star\|^2}] \leq \|S_k(A)\| \cdot \|x_0 - x_\star\|^2 + \rho^2 \gamma^2 \left( \sum_{l=1}^{k-1} \|S_{l,k}(A)\| + 1 \right) \cdot \sup \|a_{ij}\|^2$$

$$\leq \|S_k(A)\| \cdot \|x_0 - x_\star\|^2 + \rho^2 \gamma^2 \mu \sum_{l=1}^{k-1} C_{k,l} \|S_{k-l}(A)\| + \rho^2 \gamma^2 \mu$$

(4.12)

The above equality (4.12) comes from the change of index by $k - i$ and the identity $C_{k,k-i} = C_{k,i}$. Next adapting (4.8) and (4.10), we have

$$E_{\omega_0}[^{\|x_k - x_\star\|^2}] \leq \varphi^k \left( 1 + \Delta_k \right) \eta + \rho^2 \gamma^2 \mu \sum_{i=1}^{k-1} C_{k,k-i} \|S_i(A)\| + \rho^2 \gamma^2 \mu$$

(4.12)
Lemma 4.4. The following holds with constant $C$ related to $n$, $k$ and $\varphi$,

\[(4.13) \sum_{i=1}^{k-1} \exp\left(\frac{i}{n-k}\right)\varphi^i(1+\Delta_i) \leq \frac{1}{1 - \varphi \exp\left(\frac{1}{n-k}\right)} + c\varphi \exp\left(\frac{1}{n-k}\right)\]

Proof. Denote $a := \frac{1}{n-k} + \ln \varphi$, the left side of (4.13) via changing the base of the power

\[
\sum_{i=0}^{k-1} \exp\left(\frac{i}{n-k}\right)\varphi^i(1+\Delta_i) = \sum_{i=0}^{k-1} \exp\left(\frac{i}{n-k}\right)\exp(i \ln \varphi)(1+\Delta_i) = \sum_{i=0}^{k-1} \exp(ai)(1+\Delta_i)
\]

\[
= \sum_{i=0}^{k-1} \exp(ai) + \sum_{i=0}^{k-1} \exp(ai) \Delta_i = \frac{1 - \exp(ka)}{1 - \exp(a)} + \frac{1}{2n} \sum_{i=1}^{k-1} \exp(ai)i(i-1)
\]

Denote $g(a) = \sum_{i=0}^{k-1} \exp(ai)i(i-1)$, $F(x) = (x^2 - 2x + 2)e^x$, $c_a = \frac{a^2 - 2a + 2}{a^3}$

\[
g(a) \leq \int_{1}^{k} e^{ax}x^2dx = -\frac{1}{a^3}(F(a) - F(ak)) \leq \frac{F(a)}{(-a)^3} \leq c_a \varphi \exp\left(\frac{1}{n-k}\right)
\]

Therefore

\[
\sum_{i=0}^{k-1} \exp\left(\frac{i}{n-k}\right)\varphi^i(1+\Delta_i) \leq \frac{1}{1 - \varphi \exp\left(\frac{1}{n-k}\right)} + c\varphi \exp\left(\frac{1}{n-k}\right) \quad \blacksquare
\]

Proof of Theorem 4.3. Combined with the above inequality we have

\[
\mathbb{E}_{wo}[\|x_k - x_s\|] \leq \varphi^k(1+\Delta_k)\eta + \rho^2\gamma^2\mu\left(\frac{1}{1 - \varphi \exp\left(\frac{1}{n-k}\right)} + c\varphi \exp\left(\frac{1}{n-k}\right)\right) + \rho^2\gamma^2\mu
\]

\[
\leq \varphi^k\left(1 + \frac{k(k-1)}{2n}(1+C)\right)\eta + \rho^2\gamma^2\mu\left(\frac{1}{1 - \varphi \exp\left(\frac{1}{n-k}\right)} + c\varphi \exp\left(\frac{1}{n-k}\right)\right) + \rho^2\gamma^2\mu
\]

Therefore the target estimate is under control of $\varphi$, $\rho$ and $\gamma$, i.e.

\[
(4.14) \mathbb{E}_{wo}[\|x_k - x_s\|] \leq \varphi^k\left(1 + \frac{k(k-1)}{2n}(1+C)\right)\eta + \rho^2\gamma^2\mu\left(\frac{1}{1 - \varphi \exp\left(\frac{1}{n-k}\right)} + c\varphi \exp\left(\frac{1}{n-k}\right) + 1\right) \quad \blacksquare
\]

Remark 4.5. In order to make sure the inequality (4.14) hold, we need to guarantee that the radio $\varphi \in (0, 1)$. By the definition of $\varphi$, we will have

\[
(4.15) 0 < 1 - 2\gamma \sigma + \gamma^2 \sigma \mu < 1.
\]

If $\gamma < \frac{\mu}{2}$, we will have $\varphi < 1$ and $C = \frac{1}{\varphi}$ as well. Moreover, if $\sigma < \mu$, then $\gamma > 0$ exists.

Remark 4.6. The suggested strategy to win an average convergent rate: Given $\delta > 0$, then we can find a pair $(k_0, \eta_0)$, s.t $k > k_0$ and $\eta < \eta_0$, then

\[
\mathbb{E}_{wo}[\|x_k - x_s\|] \leq \delta.
\]
Denote the terms in (4.14)
\[ E_{wo}[\|x_k - x\|_2^2] \leq \phi^k \left( 1 + \frac{k(k-1)}{2n} (1 + C) \right) \eta \]
\[ + \rho^2 \gamma^2 \mu \left( \frac{1}{\phi \exp\left( \frac{1}{n-k} \right)} + c_a \phi \exp\left( \frac{1}{n-k} \right) + 1 \right) \]

- The terms in the (I) can be bounded by \( (I) \leq \phi^k \beta(k) < \delta \), if \( k < n^{1/3} \).
- The terms in the bracket of (II) can be uniformly controlled by some constant \( C \). Therefore if we want to achieve \( (II) \leq C \rho^2 \gamma^2 \mu < \delta^2 \), we just need to require \( \gamma \) small enough.

We can run the regular IGM algorithm with replacement sampling of the regular test vectors \( a_i \) as long as \( k < n^{1/3} \). If this is not the case, then we enlarge the test vectors \( a_{li} = a_i, l = 1, \cdots, m \). So the total number of the sampling pool becomes \( nm \) where \( m \geq \frac{k^3}{n} \). Therefore, we can run the IGM with new test vectors \( a_{li} \).

### 4.4. Examples satifying the assumptions of the main result.

**Example 4.7.** Let \( G \) denote a compact group and \( \pi : G \to U_d \) be an affine representation that maps an element in \( G \) to the orthogonal group \( U_d \), s.t. \( \pi(g) \pi(h) = e^{i\phi(g,h)} \pi(gh) \). An affine representation is isotropic if
\[ \int_{G} \pi(g)x \pi(g)^* dg = \frac{tr(x)}{d} \mathbb{1}_d \]
Let \( |G| = n \). Then \( \exists h \in \mathbb{C}^d, \|h\| = \sqrt{d} \)
(4.16)
\[ \frac{1}{|G|} \sum_g (gh)\langle hg^{-1} \rangle = \frac{tr(h)}{d} \mathbb{1}_d \]

1. Let \( a_g = gh\langle e_0 \rangle \) here \( e_0 \) is an unital element in \( \mathbb{C}^d \). Observe that \( \sigma = 1, \mu = d, \varphi = 1 - 2\gamma + \gamma^2 d, C = \frac{1}{\varphi} \). Then (4.9) will hold in the domain of \( \gamma \in (0, \frac{2}{d}) \) in (4.15).
2. Let \( a_g = gh\langle gh^{-1} \rangle \). Then \( \sigma = d, \mu = d^2, \varphi = 1 - 2\gamma d + \gamma d^3, C = \frac{1}{\varphi} \). Then (4.9) will hold in the domain of \( \gamma \in (0, \frac{2}{d^2}) \) in (4.15).

Spherical designs were defined by Delsarte-Goethals-Seidel in 1977 [3]. We consider a finite subset \( X \) on the unit sphere \( \mathbb{S}^{n-1} \) in n-dimensional Euclidean space \( \mathbb{R}^n \). Let \( t \) be a natural number. A finite subset \( X \subset \mathbb{S}^{n-1} \) is called a spherical t-design if
(4.17)
\[ \frac{1}{|\mathbb{S}^{n-1}|} \int_{x \in \mathbb{S}^{n-1}} f(x) d\sigma(x) = \frac{1}{|X|} \sum_{u \in X} f(u) \]
holds for any polynomial \( f(x) = f(x_1, x_2, \cdots, x_n) \) of degree at most \( t \), with the usual integral on the unit sphere.

**Example 4.8.** For 2-designs we obtain
(4.18)
\[ \int_{\mathbb{S}^{d-1}} p(x) d\sigma(x) = \frac{Id}{d} \]
Therefore taking $p(x) = |x⟩⟨x|$, by (4.17) we have
\[
\frac{1}{M} \sum_{i=1}^{M} x_i ⟨x_i | = \frac{I_d}{d}
\]
Let $x_i = \sqrt{w_i} |φ_i⟩⟨φ_i |$ with unitary $φ_i$. Then $\exists h \in S^d, \sup |⟨h, φ_i |^2 \leq 1 − \epsilon$
\[
\frac{∥h∥^2}{d} = \frac{1}{M} \sum_{i} w_i |⟨h, φ_i |^2 \leq \sup w_i \cdot (1 − \epsilon) < \sup w_i
\]
$\sup ∥x_i∥ = \sup w_i$. Then we know $σ = \frac{1}{d}, \mu = \sup w_i, \varphi = 1 − 2\frac{2}{d} + \frac{\gamma^2}{d} \sup w_i, C = \frac{1}{\varphi}$. Then (4.9) will hold in the domain of $γ \in (0, \frac{2}{\sup w_i})$ in (4.15).

References

[1] W. Albar, M. Junge, and M. Zhao. Noncommutative versions of the arithmetic-geometric mean inequality. ArXiv e-prints, March 2017.
[2] George E Andrews. The theory of partitions, volume 2. Cambridge university press, 1998.
[3] Philippe Delsarte, Jean-Marie Goethals, and Johan Jacob Seidel. Spherical codes and designs. Geometriae Dedicata, 6(3):363–388, 1977.
[4] Etienne Emmrich. Discrete versions of gronwall’s lemma and their application to the numerical analysis of parabolic problems. 1999.
[5] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. Inequalities. Cambridge university press, 1952.
[6] Arie Israel, Felix Krahmer, and Rachel Ward. An arithmetic–geometric mean inequality for products of three matrices. Linear Algebra and its Applications, 488:1–12, 2016.
[7] Marius Junge. Operator spaces and araki-woods factors: a quantum probabilistic approach. International Mathematics Research Papers, 2006, 2006.
[8] J Pecaric, Jiajin Wen, Wan-lan Wang, and T Lu. A generalization of maclaurin’s inequalities and its applications. Mathematical Inequalities and Applications, 8(4):583, 2005.
[9] Gilles Pisier et al. An inequality for $p$-orthogonal sums in non-commutative $L_p$\{p\}. Illinois Journal of Mathematics, 44(4):901–923, 2000.
[10] Benjamin Recht and Christopher Ré. Beneath the valley of the noncommutative arithmetic-geometric mean inequality: conjectures, case-studies, and consequences. arXiv preprint arXiv:1202.4184, 2012.
[11] Gian-Carlo Rota. On the foundations of combinatorial theory i. theory of möbius functions. Probability theory and related fields, 2(4):340–368, 1964.
[12] Dan Voiculescu. Lectures on free probability. Lectures Notes in Mathematics, 1738, 2000.