The Paulsen Problem Made Simple

Linus Hamilton[1]

Massachusetts Institute of Technology, 77 Massachusetts Ave, USA
luh@mit.edu

Ankur Moitra[2]

Massachusetts Institute of Technology, 77 Massachusetts Ave, USA
moitra@mit.edu

—— **Abstract** ——

The Paulsen problem is a basic problem in operator theory that was resolved in a recent tour-de-force work of Kwok, Lau, Lee and Ramachandran. In particular, they showed that every ϵ-nearly equal norm Parseval frame in d dimensions is within squared distance $O(\epsilon d^{13/2})$ of an equal norm Parseval frame. We give a dramatically simpler proof based on the notion of radial isotropic position, and along the way show an improved bound of $O(\epsilon d^2)$.

1   Introduction

The Paulsen problem is a basic problem in operator theory that was resolved in a recent work of Kwok, Lau, Lee and Ramachandran [12]. To state the problem, we need the following definition:

▶ **Definition 1.** We say that a set of vectors $v_1, v_2, \ldots, v_n \in \mathbb{R}^d$ is an *equal norm Parseval frame* if

$$\sum_{i=1}^{n} v_i v_i^T = I \text{ and } \|v_i\|_2^2 = \frac{d}{n} \text{ for each } i.$$

Alternatively, we say that it is an ϵ-*nearly equal norm Parseval frame* if

$$(1-\epsilon)I \preceq \sum_{i=1}^{n} v_i v_i^T \preceq (1+\epsilon)I \text{ and } (1-\epsilon)\frac{d}{n} \leq \|v_i\|_2^2 \leq (1+\epsilon)\frac{d}{n} \text{ for each } i.$$

When we drop the condition on the norm of each vector, we refer to the set of vectors as a *Parseval frame* or an ϵ-*nearly Parseval frame* respectively. Let $\mathcal{F}$ denote the set of

---

10th Innovations in Theoretical Computer Science (ITCS 2019).
Editor: Avrim Blum; Article No. 41; pp. 41:1–41:6
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

all equal norm Parseval frames. Lastly for two sequences of vectors $V = v_1, v_2, \ldots, v_n$ and $W = w_1, w_2, \ldots, w_n$ of the same length, we let

$$\text{dist}^2(V, W) = \sum_{i=1}^{n} \|v_i - w_i\|^2.$$

With this terminology in hand, the Paulsen problem asks:

▶ **Conjecture 2.** *For every ϵ-nearly equal norm Parseval frame V, is*

$$\inf_{W \in \mathcal{F}} \text{dist}^2(V, W)$$

bounded by a fixed polynomial in ϵ and d?

See [12] and references therein for a detailed account of the history of the Paulsen problem along with earlier bounds on the squared distance that were polynomial in ϵ, d and n. Through a tour-de-force utilizing operator scaling, connections to dynamical systems and ideas from smoothed analysis, Kwok, Lau, Lee and Ramachandran [12] proved that the squared distance is at most $O(\epsilon d^{13/2})$. The paper was 104 pages long and highly complex. Our main result is a dramatically simpler proof of the Paulsen conjecture, that also yields a much better bound:

▶ **Theorem 3** (Main). *For any ϵ-nearly equal norm Parseval frame V, there is an equal norm Parseval frame W with*

$$\text{dist}^2(V, W) \le 20\epsilon d^2.$$

In terms of lower bounds, Cahill and Casazza [4] gave a family of examples of ϵ-nearly equal norm Parseval frames where the squared distance to the closest equal norm Parseval frame is at least $\Omega(\epsilon d)$. It is an interesting open question to close this gap.

Our main idea is to make use of the notion of *radial isotropic position*[3]. In the next section, we define it formally. But to understand it informally, it is useful to compare it to the more familiar notion of placing a set of vectors in isotropic position: Given a set of vectors $V = v_1, v_2, \ldots, v_n \in \mathbb{R}^d$, is there an invertible affine transformation that generates a new set of vectors $Y = Av_1 + b, Av_2 + b, \ldots, Av_n + b$ that has mean zero and identity covariance? It is well known that there is such a transformation if and only if $\sum_i v_i v_i^T$ has full rank.

However such a transformation can also stretch out some directions much more than others – e.g. if all but one of the vectors in V are contained in a $d-1$-dimensional subspace. In this case, the set of vectors after applying the transformation would be quite far from where it started out, in total squared distance. Informally, radial isotropic position asks for a linear transformation A so that the *renormalized* vectors $w_i = Av_i / \|Av_i\|$ have the property that $\sum_i w_i w_i^T$ is a scalar multiple of the identity. The transformation is now nonlinear but is particularly well suited for constructing a close by equal norm Parseval frame.

One can now ask the same sort of question as before: When can a set of vectors be placed in radial isotropic position? Barthe [2] gave a complete characterization of when this is and is not possible which in turn plays a key role in our proof. It turns out that a sufficient condition is that every d vectors are linearly independent. Now we construct an equal norm Parseval frame as follows: First we renormalize the vectors in V and then we perturb them.

---

[3] This concept goes by many other names in the literature, such as *well-spread vectors* [5] or *geometric scaling for rank one Brascamp-Lieb datum* [9]. The name we use here originated in [11].

Perturbations play a delicate role in [12]. They give a dynamical system which constructs an equal norm Parseval frame from an ϵ-nearly equal norm Parseval frame as its input. In order to bound the total squared distance between the input and output, they need to lower bound the convergence rate. They do this through a certain pseudorandom property (Definition 4.3.2) which they show holds when the input is appropriately perturbed. In our proof, all we need is that the perturbations do not move the set of points by too much in squared distance and that afterwards every d of them are linearly independent[4]. The latter condition guarantees that there is a linear transformation that places them in radial isotropic position. Let W be the set of vectors, after applying the linear transformation and renormalizing. By definition, it is an equal norm Parseval frame. Our main technical contribution is in bounding the squared distance between V and W, which we do through some elementary but subtle algebraic manipulations.

Taking a step back, the notion of radial isotropic position seems quite powerful and mysterious but has thus far only found a handful of applications. Forster [7] used it to prove a remarkable lower bound in communication complexity (by lower bounding the sign rank of the Hadamard matrix). Hardt and Moitra [11] gave the first algorithm for computing the transformation that places a set of vectors in radial isotropic position (under a slight strengthening of Barthe's conditions). They also gave applications to linear regression in the presence of outliers. Dvir, Saraf and Wigderson [5] used it to prove superquadratic lower bounds for 3-query locally correctable codes over the reals. Here we use it to give a simple proof of the Paulsen conjecture. Are there other exciting applications waiting to be discovered?

Connections to Operator Scaling and the Brascamp-Lieb Inequality

Radial isotropic position is itself a special case of the more general notion of *geometric position* [1, 2] where we are given an n tuple of linear transformations $B_1, B_2, \ldots, B_n$ of dimensions $d_1 \times d, d_2 \times d, \ldots, d_n \times d$ and a nonnegative vector c of dimension n with $\sum_{i=1}^{n} c_i d_i = d$ and the goal is to find square, invertible matrices $A_1, A_2, \ldots, A_n$ and A so that

$$\sum_{i=1}^{n} c_i \left(A_i^{-1} B_i A \right)^T \left(A_i^{-1} B_i A \right) = I \text{ and } \left(A_i^{-1} B_i A \right) \left(A_i^{-1} B_i A \right)^T = I \text{ for each } i.$$

If we set $d_i = 1$ for all i, then each linear transformation B_i can be written as the inner-product with some vector v_i. Now if we also set $c_i = \frac{d}{n}$ for all i, it is easy to check that A places the set of vectors $v_1, v_2, \ldots, v_n$ in radial isotropic position.

It turns out that having $A_1, A_2, \ldots, A_n$ and A that place $B_1, B_2, \ldots, B_n$ in geometric position with respect to the vector c yields an explicit expression for the best constant C for which the inequality

$$\int_{x \in \mathbb{R}^d} \prod_{i=1}^{n} \left(f_i(B_i x) \right)^{c_i} dx \leq C \prod_{i=1}^{n} \left(\int_{x_i \in \mathbb{R}^{d_i}} f_i(x_i) dx_i \right)^{c_i}$$

holds over all m tuples of nonnegative functions $f_1, f_2, \ldots, f_m$ [3]. This is called the Brascamp-Lieb inequality.

---

[4] In particular, essentially all sufficiently small perturbations would work for us. It could even be an infinitesimal perturbation because we do not need any quantitative bounds on how far they are from having a non-trivial linear dependence.

Finally, in terms of how to compute $A_1, A_2, \ldots, A_n$ and A, a popular approach is *operator scaling* [10] and there has been considerable recent progress in bounding the number of iterations it needs [8, 9]. As we mentioned, Kwok, Lau, Lee and Ramachandran [12] used operator scaling to solve the Paulsen conjecture. In this sense, our approach and theirs are closely related in that they both revolve around algorithms (in our case the ellipsoid algorithm) for computing radial isotropic position. Perhaps the main technical divergence is that they track how the squared distance changes after each iteration of operator scaling, while we are able to bound the squared distance just based on transformation that places $v_1, v_2, \ldots, v_n$ into radial isotropic position. It is also worth mentioning that if instead of proving existence of a nearby equal norm Parseval frame, we want to find it up to some target precision δ, the approaches based on operator scaling typically require the number of iterations to be polynomial in $1/\delta$. In contrast, we will give algorithms whose running time is polynomial in $\log 1/\delta$.

2 Radial Isotropic Position and the Proof

First we introduce some of the basic concepts and results about radial isotropic position. We will do so in slightly more generality than we will ultimately need.

▶ **Definition 4.** We say that a set of vectors $u_1, u_2, \ldots, u_n \in \mathbb{R}^d$ is in *radial isotropic position* with respect to a coefficient vector $c \in \mathbb{R}^n$ if

$$\sum_{i=1}^{n} c_i \left(\frac{u_i}{\|u_i\|} \right) \left(\frac{u_i}{\|u_i\|} \right)^T = I.$$

Note that if we take the trace of both sides in the expression, we get the necessary condition that $\sum_{i=1}^{n} c_i = d$. In fact we will only ever consider the case when each $c_i = \frac{d}{n}$. We will also need the following key definition:

▶ **Definition 5** ([6]). For a set U of vectors $u_1, u_2, \ldots, u_n \in \mathbb{R}^d$, its *basis polytope* is defined as

$$\mathcal{B}(U) = \left\{ c \in \mathbb{R}^n \text{ s.t. } \sum_{i=1}^{n} c_i = d \text{ and for all } A \subseteq [n], \dim \left(\operatorname{span}\{u_i\}_{i \in A} \right) \geq \sum_{i \in A} c_i \right\}.$$

Now we are ready to state Barthe's theorem:

▶ **Theorem 6** ([2]). *A set of vectors $U = u_1, u_2, \ldots, u_n \in \mathbb{R}^d$ can be put into radial isotropic position with respect to c by a linear transformation if and only if $c \in \mathcal{B}(U)$.*

Some further remarks: (1) The usual definition of the basis polytope is based on taking the convex hull of the indicators of subsets of vectors in U that form a basis. (2) The alternative definition we gave in Definition 5 will be more directly useful for our purposes, and was proven to be equivalent by Edmonds [6]. He used this equivalence to give a separation oracle for the basis polytope, which in turn plays a key role in the algorithm of Hardt and Moitra [11] for computing the linear transformation that puts a set of vectors into radial isotropic position.

See `https://arxiv.org/abs/1809.04726` for the proof of the main theorem. The main idea is to renormalize and then perturb the vectors in V. After the perturbation, we can invoke Theorem 6 to find a transformation A that places the vectors in radial isotropic position. We show that we can assume without loss of generality that A is a nonnegative diagonal matrix whose entries are sorted in non-increasing order along the diagonal. Our main technical lemma gives a bound on the squared distance:

▶ **Lemma 7.** *Suppose U is a 4ϵ-nearly Parseval frame and A is an entrywise diagonal matrix that puts U in radial isotropic position. Also suppose that for each i*

$$(1 - \gamma')\frac{d}{n} \leq \|u_i\|_2^2 \leq (1 + \gamma')\frac{d}{n}.$$

Now set

$$W = w_1, w_2, \ldots, w_n \ with \ w_i \triangleq \sqrt{\frac{d}{n}}\Big(\frac{Au_i}{\|Au_i\|}\Big).$$

Then we have that $\mathrm{dist}^2(U, W) \leq 8\epsilon d^2 + 4\gamma'd^2$.

See `https://arxiv.org/abs/1809.04726` for the proof of the main technical lemma.

3 An Algorithm for the Paulsen Problem

Every step of the proof of Theorem 3 is straightforward to implement algorithmically, except for the step where we compute the transformation A that places the set of vectors U in radial isotropic position. Fortunately, Hardt and Moitra [11] gave an algorithm for computing A under a slight strengthening of Barthe's conditions which holds in our setting. Informally, they require the vector c to be strictly inside the basis polytope according to the following notion of scaling:

▶ **Definition 8.** *Let $(1 - \alpha)\mathcal{B}(U)$ denote the set of vectors c with the following properties: (1) $\sum_{i=1}^n c_i = d$, (2) $0 \leq c_i \leq 1$ for all i and (3) for all nonnegative directions u with $u_{min} = 0$,*

$$(1 - \alpha)\max_{v \in \mathcal{B}(U)} u^T v \geq u^T c.$$

We will state a special case of their main theorem, which is sufficient for our purposes.

▶ **Theorem 9** ([11]). *Let $\delta > 0$ and $\alpha > 0$. Suppose $U = u_1, u_2, \ldots, u_n \in \mathbb{R}^d$ has the property that every set of d vectors are linearly independent. Then given $c \in (1 - \alpha)\mathcal{B}(U)$, there is an algorithm to find a linear transformation A so that*

$$\sum_{i=1}^n c_i\Big(\frac{Au_i}{\|Au_i\|}\Big)\Big(\frac{Au_i}{\|Au_i\|}\Big)^T = I + J$$

where $\|J\|_\infty \leq \delta$ – i.e. the largest entry of J in absolute value is at most δ. The running time is polynomial in $1/\alpha$, $\log 1/\delta$ and L where L is an upper bound on the bit complexity of U and c.

By combining their algorithm with our proof of Theorem 3 we get:

▶ **Corollary 10.** *Suppose $V = v_1, v_2, \ldots, v_n \in \mathbb{R}^d$ is an ϵ-nearly equal norm Parseval frame. Furthermore suppose $n > d$. Then given $\delta > 0$, there is an algorithm to compute a δ-nearly equal norm Parseval frame W with*

$$\mathrm{dist}^2(V, W) \leq 20\epsilon d^2$$

whose running time is polynomial in $\log 1/\delta$ and L where L is an upper bound on the bit complexity of V.

See `https://arxiv.org/abs/1809.04726` for the proof of the corollary. This answers an open question of [12], where they ask whether there is an algorithm for finding an equal norm Parseval frame up to some precision δ whose running time is polynomial in $\log 1/\delta$.

────── **References** ──────

1 Keith Ball. Volumes of sections of cubes and related problems. In *Geometric aspects of functional analysis*, pages 251–260. Springer, 1989.

2 Franck Barthe. On a reverse form of the Brascamp-Lieb inequality. *Inventiones mathematicae*, 134(2):335–361, 1998.

3 Jonathan Bennett, Anthony Carbery, Michael Christ, and Terence Tao. The Brascamp–Lieb inequalities: finiteness, structure and extremals. *Geometric and Functional Analysis*, 17(5):1343–1415, 2008.

4 Jameson Cahill and Peter G Casazza. The Paulsen problem in operator theory. *submitted to Operators and Matrices*, 2011.

5 Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Superquadratic Lower Bound for 3-Query Locally Correctable Codes over the Reals. *Theory of Computing*, 13(1):1–36, 2017.

6 Jack Edmonds. Submodular functions, matroids, and certain polyhedra. *Combinatorial structures and their applications*, pages 69–87, 1970.

7 Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *Journal of Computer and System Sciences*, 65(4):612–625, 2002.

8 Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. A deterministic polynomial time algorithm for non-commutative rational identity testing. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 109–117. IEEE, 2016.

9 Ankit Garg, Leonid Gurvits, Rafael Oliveira, and Avi Wigderson. Algorithmic and optimization aspects of Brascamp-Lieb inequalities, via Operator Scaling. *Geometric and Functional Analysis*, 28(1):100–145, 2018.

10 Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004.

11 Moritz Hardt and Ankur Moitra. Algorithms and hardness for robust subspace recovery. In *Conference on Learning Theory*, pages 354–375, 2013.

12 Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, and Akshay Ramachandran. The Paulsen problem, continuous operator scaling, and smoothed analysis. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 182–189. ACM, 2018.