

Geometric Perspective on Concentration Phenomena in Frame Theory

Samuel A. Ballas

Ferhat Karabatman

Tom Needham

August 21, 2026

Abstract

Parseval and equal-norm frames play a fundamental role in frame theory and signal processing. It is known that a random frame, with unit vectors drawn independently from the uniform distribution on the sphere, will be nearly Parseval with high probability; asymptotic results go back at least to Goyal, Vetterli, and Thao and a non-asymptotic error bound was proved more recently by Kwok, Lau, and Ramachandran. In this work, we prove a dual result, which shows that random Parseval frames, with respect to the Haar measure, are nearly equal-norm with high probability. Our proofs are geometric in nature, and rely on general measure concentration principles in Riemannian manifolds. Using these techniques, we also give a novel probabilistic upper bound for the Paulsen problem.

1 Introduction

Frames provide redundant systems of vectors for representing signals in Hilbert spaces. Unlike bases, frames allow stable reconstruction while permitting redundancy, a feature that makes them particularly useful in signal processing, data transmission, and related areas [6, 10, 11, 21]. In finite dimensions, a frame is simply a spanning family of vectors $(x_i)_{i=1}^n \subset \mathbb{R}^d$, and its redundancy is reflected in the fact that one may have $n > d$. This redundancy is not merely a formal feature: it plays a central role in the stability of reconstruction, especially when the transmitted coefficients are affected by noise or erasures.

A central class of frames is given by *Parseval frames*. For such frames, the *frame operator*, or the map $\mathbb{R}^d \rightarrow \mathbb{R}^d$ defined by

$$v \mapsto \sum_{i=1}^n \langle v, x_i \rangle x_i,$$

is the identity map. The interpretation is that, in this case, the signal v can be reconstructed from its sequence of linear measurements $(\langle v, x_i \rangle)_{i=1}^n \in \mathbb{R}^n$ in a particularly simple and stable manner. Equal-norm tight frames, where $\|x_i\| = \|x_j\|$ for all i, j , are also especially important because they distribute the importance of the linear measurements uniformly. This balance is closely connected to robustness under coefficient losses and to optimality properties in reconstruction problems. Several foundational results illustrate the importance of these structures. Goyal et al. showed that, under an additive noise model, the mean squared error is minimized precisely by tight frames [16]. Subsequent works further emphasized the role of tight and equal-norm tight frames in

reconstruction problems involving noise and erasures. For instance, Casazza and Kovačević [9] and Holmes and Paulsen [20] showed that equal-norm tight frames enjoy strong optimality properties in erasure settings. Taken together, these results indicate that both tightness and equal distribution of vector norms are essential features for stable and robust signal representation.

In light of the above, frames which are both Parseval and equal-norm are desirable in applications. Unfortunately, the space of equal-norm Parseval frames forms a complicated subvariety of $\mathbb{R}^{d \times n}$ which is hard to sample directly [1, 5, 12, 28, 30]. On the other hand, spaces of frames which are only constrained to be Parseval or to be equal-norm (i.e., only satisfying one of the desired conditions) are easy to sample, as they can be identified with Stiefel manifolds or products of spheres, respectively. The goal of this paper is to address the sampling problem from a measure concentration perspective, addressing the natural question: if one samples vectors satisfying one of the desired conditions (e.g., the Parseval condition), how likely is it that the resulting system is close to satisfying the other condition (e.g., the unit norm condition)? In other words, to what extent are the frame-theoretic structures that are optimal for reconstruction typical within natural random ensembles?

1.1 Contributions and Outline

In this work, we address the questions above from the perspective of measure concentration on Riemannian manifolds (see [25, 36]). An outline of the paper and descriptions of our main contributions are as follows:

Concentration on Parseval Frames. In Section 2, we consider random equal-norm frames, considered as i.i.d. samples from the uniform measure on a sphere. It is known in the frame theory community that random equal norm frames (i.e., sample from the sphere) are highly likely to be nearly Parseval. Different versions of this statement are proved in [3, 17, 23]; see Section 1.2 for more details on related work. In Theorem 2.1, we prove a version of this statement which originally appeared in [23]: random unit norm frames are close to being Parseval with high probability, when the number of vectors in the frame is large—in fact, we arrived at this proof independently before being made aware of its existence in the literature, and the proof is included here for completeness and because it will be useful later in the paper. The proof relies on classical matrix concentration inequalities. Similar techniques allow us to establish concentration on Parseval frames for frames whose vector norms are not fully prescribed, but only bounded above (Theorem 2.4). The tightness of these concentration bounds is explored via numerical experiments, included in the appendix.

Concentration on Equal-Norm Frames. We take the dual perspective in Section 3, and consider concentration properties of random Parseval frames, considered as samples from a symmetric measure on the Stiefel manifold. We show in Theorem 3.1 that these frames are nearly equal-norm with high probability, assuming that the number of frame vectors is sufficiently large. The proof is more technical than the sphere case, and requires some estimates for geodesic distance and Ricci curvature of the Stiefel manifold. We also include numerical experiments illustrating this concentration result in the appendix.

Application to the Paulsen Problem. We give an application of our sphere concentration result to the Paulsen problem [3, 4, 7, 18, 22]. Roughly speaking, the Paulsen problem seeks a uniform bound on the squared distance from a frame that is ε -nearly equal-norm and ε -nearly Parseval (in a certain precise sense)—referred to briefly as an ε -nearly equal-norm Parseval frame—to the

subvariety of *exactly* equal-norm Parseval frames. In Theorem 4.3, we derive an $O(\varepsilon^2 d)$ bound which holds with high probability for random nearly equal-norm and Parseval frames. Our bound is tighter than any of the deterministic bounds in the existing literature. In fact, we show in Proposition 4.8 that any deterministic bound can be at best on the order of $O(\varepsilon d)$; this claim has been made in the literature, but to our knowledge no formal proof has appeared previously. This shows, in particular, that an optimal deterministic bound and our probabilistic bound are of different orders. Our proof is based on an extension of an example appearing in [32]. We note that similar probabilistic Parseval problem bounds, using a different sampling scheme, have appeared previously in [23, 32]; the distinction of our result is explained in the following subsection.

1.2 Related Work

Random frames have been studied in connection with quantization, erasures, uncertainty principles, and probabilistic constructions; see, e.g., [2, 14, 17, 26, 35]. For instance, a basic model for random frames considers them as rectangular matrices with i.i.d. Gaussian entries. It follows from general measure concentration results that such a frame is highly likely to be nearly equal norm and nearly Parseval (after appropriate normalizations)—see [36, Theorem 3.1.1 and 4.6.1], or related estimates in [2, Corollaries 3.2 and 3.5]. More generally, Lyubarskii and Vershynin [26], prove high-probability uncertainty principle estimates for random frame models, including subgaussian matrices, with applications to Kashin representations and vector quantization.

This paper is primarily concerned with random matrices sampled from distributions with additional structure imposed. There are several results in the literature which work with random frames whose vectors are sampled independently from a sphere, i.e., random unit-norm frames. Early work of Goyal, Vetterli, and Thao shows that the resulting frames are tight, asymptotically in the number of frame vectors [17, Theorem 1]. In the non-asymptotic regime, the earliest result we are aware of is due to Bodmann, who gives very general probabilistic guarantees on near-tightness for random fusion frames sampled from unitarily invariant distributions [2, Theorem 3.6]; this specializes in the rank-1 case to essentially the setting of random unit norm frames. By working specifically with the case of unit norm frames, Kwok, Lau, and Ramachandran are able to derive much tighter probabilistic results [23]. Theorem 2.1 below reproduces a result which can be found in the proof of [23, Lemma 5.4], giving a non-asymptotic bound on the probability that a random unit norm frame is nearly Parseval. We include a proof of this result for the sake of completeness, and because Theorem 2.1 will be used in a crucial way to prove Theorem 4.3 later in the paper. Finally, we note that a stronger concentration result of this form is given in [15, Lemma A.13], following from [35, Theorem 5.39], but these results involve unspecified constants. We prefer to work with fully explicit concentration bounds for the random models considered here, which allows one to examine the sharpness of these bounds through Monte Carlo simulations.

Rather than imposing the equal norm condition and studying the probability that the resulting frame is nearly Parseval, Theorem 3.1 works in the opposite direction: we sample random Parseval frames by identifying the space of frames with the Stiefel manifold, endowed with Haar measure, and study the probability that the resulting frame is nearly equal norm. A related concentration phenomenon for random Parseval frames can also be inferred from the results of Hayden, Leung, and Winter [19]. In particular, their concentration estimates for random projections imply, under the standard correspondence with Parseval frames, that the norm of each individual frame vector concentrates around the common equal-norm value. Our perspective is different: rather than

controlling the deviation of the frame vectors from this value one at a time, we consider a frame chosen randomly from the space of Parseval frames and study its distance, as a whole, from the equal-norm condition. Thus, our result provides a global concentration statement for the frame rather than a vector-wise concentration estimate. This is in contrast to the more standard approach of deriving concentration results for a single frame then extending to the full frame via a union bound argument; see Remark 3.11 for further discussion of the distinction.

The connection with the Paulsen problem was also considered in [23]. In that work, the authors obtain a high-probability $O(\varepsilon^2 d)$ -type upper bound for the Paulsen distance in a random equal-norm model. Ramachandran later refined related results in his dissertation [32], using the result of [15]. These results are closely related to ours, but the underlying sampling models are different. The Paulsen problem concerns frames which are simultaneously ε -nearly equal-norm and ε -nearly Parseval. By contrast, the random frames considered in the works above are exactly equal-norm, rather than nearly equal-norm. Thus, the randomness is imposed on a smaller and more rigid class of frames than the one appearing naturally in the Paulsen problem. Our result addresses this distinction directly. We sample from the space of frames satisfying the two approximate Paulsen conditions themselves, and show that a high-probability $O(\varepsilon^2 d)$ upper bound holds on this natural space. Moreover, our argument avoids the additional pseudorandomness and operator-scaling machinery used in some of the earlier works. Instead, we use a direct geometric approach: radial projection to the sphere, the explicit sphere concentration estimate from Theorem 2.1, the deterministic Hamilton–Moitra bound [18], and the triangle inequality.

Ramachandran’s dissertation [32] also contains an $O(\varepsilon d)$ upper bound for certain frames in a restricted range of parameters. This does not contradict the high-probability $O(\varepsilon^2 d)$ estimate proved here. Rather, the two results apply in different regimes: the $O(\varepsilon d)$ estimate is relevant when n is not too large relative to d , whereas our $O(\varepsilon^2 d)$ estimate holds with high probability once n is sufficiently large. To support the optimality of the $O(\varepsilon d)$ behavior in the former regime, Ramachandran uses the $d = 2, n = 4$ example appearing in Equation (38) of [8]. In Proposition 4.8, we extend this construction to the regime $n = 2d$, showing that deterministic Paulsen bounds can be no better than order $\Omega(\varepsilon d)$ in general.

Acknowledgements

We would like to thank Akshay Ramachandran for providing detailed feedback earlier versions of the paper. TN was supported by NSF grants DMS 2324962 and CIF 2526630.

2 Concentration of Parseval Frames

In this section, we establish concentration results for random vectors drawn from appropriately scaled spheres and balls. Our main results show that random frames drawn from these spaces are close to being Parseval with high probability, with respect to both the operator and Frobenius norms. We begin with a review of basic terminology and notation from frame theory.

2.1 Frame Theoretic Concepts and Notation

Throughout the paper, we consider $\mathbb{R}^d$ as an inner product space, endowed with its standard inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. The adjoint (i.e., transpose) of a linear map $T : \mathbb{R}^d \rightarrow \mathbb{R}^n$ with respect to the standard inner products will be denoted by $T^* : \mathbb{R}^n \rightarrow \mathbb{R}^d$.

A finite sequence $(x_i)_{i=1}^n$ of vectors $x_i \in \mathbb{R}^d$ is called a *frame* if it forms a spanning set for $\mathbb{R}^d$. Note that, for the sake of convenience, we consider a frame to be an ordered tuple of vectors (rather than an unordered set, which is an alternative convention).

For any finite sequence of vectors $(x_i)_{i=1}^n$, we have the associated *analysis operator* $T : \mathbb{R}^d \rightarrow \mathbb{R}^n$, defined by $T(x) = (\langle x, x_1 \rangle, \dots, \langle x, x_n \rangle)$, and its adjoint, the *synthesis operator* $T^* : \mathbb{R}^n \rightarrow \mathbb{R}^d$, $T^*(c_1, \dots, c_n) = \sum_{i=1}^n c_i x_i$. The *frame operator* is $S = T^*T : \mathbb{R}^d \rightarrow \mathbb{R}^d$, which we write as a sum of outer products $S = \sum_{i=1}^n x_i x_i^*$. In the context of signal processing, where $x \in \mathbb{R}^d$ is considered as a signal, application of the analysis operator amounts to taking linear measurements encoded by the sequence, and application of the synthesis operator amounts to reconstructing a signal from these measurements. A simple linear algebra exercise shows that a sequence is a frame if and only if its frame operator S is invertible.

For frames in finite-dimensional spaces, there are certain natural classes of frames which are of particular interest. First, it is especially convenient when the frame operator is not just invertible, but is actually the identity map; in this case, the frame is said to be *Parseval*. Second, one may wish for the vectors in the frame to be *equal-norm*. If a frame $(x_i)_{i=1}^n$ in $\mathbb{R}^d$ is both Parseval and equal-norm, a simple calculation shows that the norm must be given by $\|x_i\|^2 = \frac{d}{n}$.

In short, the results of this paper address the following question: if one of the desired properties above is enforced (e.g., the equal-norm condition), how likely is it that a random frame is close to satisfying the other property (e.g., Parseval)?

2.2 Concentration for Spheres

We begin by considering concentration properties for equal-norm frames. We use the notation $\mathbb{S}_R^{d-1}$ for the $(d-1)$ -dimensional sphere of radius $R > 0$, considered as a submanifold of $\mathbb{R}^d$. We endow the sphere with the Riemannian metric induced from $\mathbb{R}^d$, and consider it as a measure space with normalized Riemannian volume; i.e., with the uniform probability measure.

We consider concentration properties framed in terms of operator norms. For a linear operator on Euclidean space $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$, we use $\|A\|_{\text{op}}$ to denote its operator norm and $\|A\|_F$ to denote its Frobenius norm. The main result of this section is the following.

Theorem 2.1 (Concentration over the Sphere [23]). *Let $(x_i)_{i=1}^n$ be a set of independent random vectors, drawn from the uniform distribution on $\mathbb{S}_{\sqrt{d/n}}^{d-1}$, and let S be the associated frame operator. Then, for every $\varepsilon \geq 0$,*

$$\mathbb{P}(\|S - I_d\|_{\text{op}} \geq \varepsilon) \leq 2d \exp\left(-\frac{n\varepsilon^2}{2d - 2 + \frac{2\varepsilon(d-1)}{3}}\right)$$

and

$$\mathbb{P}(\|S - I_d\|_F \geq \varepsilon) \leq 2d \exp\left(-\frac{n\varepsilon^2}{2d^2 - 2d + \frac{2\varepsilon\sqrt{d(d-1)}}{3}}\right).$$

In the language of frame theory, this result has the following interpretation: for fixed d , a random equal-norm frame consisting of n vectors is arbitrarily close to being Parseval with overwhelmingly high probability, provided n is sufficiently large. This is concretely quantified in Corollary 2.6 below. As was explained in the introduction, this result appears in the proof of [22, Lemma 5.4]. Before being made aware of its existence in the literature, we had independently found a proof, using the same main tool (the Matrix Bernstein Inequality, presented below). We provide a proof of the theorem for the sake of completeness, and because the result will be useful later on in the proof of Theorem 4.3.

Theorem 2.2 (Matrix Bernstein Inequality [34, Theorem 1.4]). *Let Y be a random self-adjoint $d \times d$ matrix such that $\mathbb{E}[Y] = 0$ and $\|Y\|_{\text{op}} \leq A$ almost surely, for some $A > 0$. Let $Y_1, \dots, Y_n$ be i.i.d. copies of Y . Then, for all $\varepsilon \geq 0$,*

$$\mathbb{P}\left(\left\|\sum_{i=1}^n Y_i\right\|_{\text{op}} \geq \varepsilon\right) \leq 2d \cdot \exp\left(\frac{-\varepsilon^2/2}{\sigma^2 + A\varepsilon/3}\right), \quad \text{where } \sigma^2 := \left\|\sum_{i=1}^n \mathbb{E}[Y_i^2]\right\|_{\text{op}}.$$

Proof of Theorem 2.1. Let $v = (v^{(1)}, \dots, v^{(d)})$ be a random vector drawn uniformly from the sphere $\mathbb{S}^{d-1}_{\sqrt{d/n}}$. By rotational invariance,

$$\mathbb{E}[v^{(i)}v^{(j)}] = 0 \quad \text{for } i \neq j, \quad \mathbb{E}[(v^{(i)})^2] = \frac{1}{n}, \quad \text{and hence } \mathbb{E}[vv^*] = \frac{1}{n}I_d.$$

Let $(x_i)_{i=1}^n$ be independent random vectors with the same distribution as v , and define the frame operator $S = \sum_{i=1}^n x_i x_i^*$. Then

$$\mathbb{E}[S] = \sum_{i=1}^n \mathbb{E}[x_i x_i^*] = n \cdot \frac{1}{n} I_d = I_d.$$

Define $Y_i := x_i x_i^* - \frac{1}{n} I_d$. Then $\mathbb{E}[Y_i] = 0$. Since $\|x_i\|^2 = d/n$, the matrix $x_i x_i^*$ has eigenvalues d/n and 0 (with multiplicity $d-1$). Hence Y_i has eigenvalues $\frac{d-1}{n}$ and $-\frac{1}{n}$ (with multiplicity $d-1$), so for $d \geq 2$,

$$\|Y_i\|_{\text{op}} = \frac{d-1}{n}.$$

Moreover, a direct computation shows that

$$\left\|\sum_{i=1}^n \mathbb{E}[Y_i^2]\right\|_{\text{op}} = \frac{d-1}{n}.$$

Applying Theorem 2.2 yields the desired bound for the operator norm.

The result in terms of the Frobenius norm follows directly from the relationship between the operator norm and the Frobenius norm:

$$\sqrt{d} \|A\|_{\text{op}} \geq \|A\|_F.$$

□

Remark 2.3. *The same approach extends beyond the uniform sphere measure. In general, one may define*

$$Y_i := x_i x_i^* - \mathbb{E}[x_i x_i^*],$$

so that $\mathbb{E}[Y_i] = 0$. Thus, the concentration result in Theorem 2.2 applies as long as the random vectors x_i satisfy $\mathbb{E}[x_i x_i^] = \alpha I_d$ for some $\alpha > 0$. In particular, this includes distributions supported on the sphere or the ball that are isotropic after appropriate normalization.*

2.3 Concentration for Euclidean Balls

The next main result is a concentration result for frames whose vectors are drawn from Euclidean balls. We use $\mathbb{B}_R^d$ to denote the closed Euclidean ball in $\mathbb{R}^d$ of radius $R > 0$. This is treated as a metric measure space with Euclidean metric and uniform probability measure. The main result of this subsection is as follows.

Theorem 2.4 (Concentration in the Ball). *Let $(x_i)_{i=1}^n$ be a set of independent random vectors, drawn from the uniform distribution on $\mathbb{B}_{\sqrt{\frac{d+2}{n}}}^d$ and let S be the associated frame operator. Then, for every $\varepsilon \geq 0$,*

$$\mathbb{P}(\|S - I_d\|_{\text{op}} \geq \varepsilon) \leq 2d \exp\left(-\frac{n\varepsilon^2}{2d + 2 + \frac{2\varepsilon(d+1)}{3}}\right)$$

and

$$\mathbb{P}(\|S - I_d\|_{\text{F}} \geq \varepsilon) \leq 2d \exp\left(-\frac{n\varepsilon^2}{2d^2 + 2d + \frac{2\varepsilon\sqrt{d}(d+1)}{3}}\right)$$

.

Sharper estimates could be derived from the methods of [15], but these are largely based on [35, Theorem 5.39], which involves unspecified constants. The proof strategy presented below yields the explicit expressions in the theorem.

Lemma 2.5. *Let v be a random vector uniformly distributed on the Euclidean ball $\mathbb{B}_r^d$. Then*

$$\mathbb{E}[vv^*] = \frac{r^2}{d+2} I_d.$$

In particular, choosing $r = \sqrt{\frac{d+2}{n}}$ ensures that $\mathbb{E}[vv^] = \frac{1}{n} I_d$.*

Proof. Since the uniform distribution on $\mathbb{B}_r^d$ is rotationally invariant, the matrix $\mathbb{E}[vv^*]$ must be a scalar multiple of the identity. In particular, all diagonal entries are equal, and we have $\mathbb{E}[v_i^2] = \frac{1}{d} \mathbb{E}[\|v\|^2]$. Let

$$k = \frac{1}{d} \text{tr}(\mathbb{E}[vv^*]) = \frac{1}{d} \mathbb{E}[\|v\|^2].$$

Then

$$\mathbb{E}[vv^*] = k I_d.$$

To compute k , note that

$$k = \frac{1}{d \text{Vol}(\mathbb{B}_r^d)} \int_{\mathbb{B}_r^d} \|v\|^2 dV.$$

Using polar coordinates (ρ, θ) , where $\|v\| = \rho$ and $dV = \rho^{d-1} d\rho d\theta$, we obtain

$$\int_{\mathbb{B}_r^d} \|v\|^2 dV = \int_0^r \int_{S^{d-1}} \rho^2 \rho^{d-1} d\theta d\rho = \left(\int_0^r \rho^{d+1} d\rho \right) \text{Vol}(S^{d-1}) = \frac{r^{d+2}}{d+2} \text{Vol}(S^{d-1}).$$

On the other hand, the volume of $\mathbb{B}_r^d$ is given by

$$\text{Vol}(\mathbb{B}_r^d) = \int_0^r \rho^{d-1} d\rho \cdot \text{Vol}(S^{d-1}) = \frac{r^d}{d} \text{Vol}(S^{d-1}).$$

Substituting these expressions yields

$$k = \frac{\frac{r^{d+2}}{d+2} \text{Vol}(S^{d-1})}{d \cdot \frac{r^d}{d} \text{Vol}(S^{d-1})} = \frac{r^2}{d+2}.$$

Consequently,

$$\mathbb{E}[vv^*] = \frac{r^2}{d+2} I_d.$$

Choosing $r = \sqrt{\frac{d+2}{n}}$ makes the expected value $\frac{1}{n} I_d$. □

Proof of Theorem 2.4. For the concentration result in the operator norm, we proceed exactly as in the corresponding part of the proof of Theorem 2.1. By Lemma 2.5, the radius of the ball must be $\sqrt{(d+2)/n}$ in order to ensure that $\mathbb{E}[S] = I_d$. For the Frobenius norm, we proceed in the same way as in Theorem 2.1. □

We conclude this section with a corollary that quantifies our concentration in terms of increasing n (the number of vectors in the frames). The proof follows immediately from the main theorems.

Corollary 2.6. *Let $\delta \in (0, 1)$ and $\varepsilon > 0$. Then, for both the sphere and ball models, there exists a constant $C_{d,\varepsilon}$ (depending on dimension d and tolerance ε) such that if*

$$n \geq C_{d,\varepsilon} \log\left(\frac{2d}{\delta}\right),$$

then

$$\mathbb{P}(\|S - I_d\|_{\text{op}} \geq \varepsilon) \leq \delta.$$

In particular:

- *To guarantee a confidence level of 95%, it suffices to take*

$$n \geq C_{d,\varepsilon} \log(40d).$$

- *To guarantee a confidence level of 99%, it suffices to take*

$$n \geq C_{d,\varepsilon} \log(200d).$$

Here one may take

$$C_{d,\varepsilon} = \frac{2d - 2 + \frac{2\varepsilon(d-1)}{3}}{\varepsilon^2}$$

for the sphere model, and

$$C_{d,\varepsilon} = \frac{2d + 2 + \frac{2\varepsilon(d+1)}{3}}{\varepsilon^2}$$

for the ball model.

An analogous statement holds for the Frobenius norm.

3 Concentration of Equal-Norm Frames

In the previous section, we showed that random collections of equal-norm vectors concentrate around Parseval frames. In this section, we prove a complementary result: random Parseval frames concentrate around equal-norm frames. More precisely, we show that if a Parseval frame is chosen at random, then the norms of its frame vectors are close to the common norm of an equal-norm Parseval frame, namely $\sqrt{d/n}$, with high probability.

A Parseval frame $(x_i)_{i=1}^n \subset \mathbb{R}^d$ can be identified with a matrix $U \in \mathbb{R}^{n \times d}$ whose rows are the vectors x_i . Such a matrix thus satisfies the condition $U^*U = I_d$, so that Parseval frames correspond to the points of the *Stiefel manifold*

$$\text{St}(n, d) := \{U \in \mathbb{R}^{n \times d} \mid U^*U = I_d\}. \quad (1)$$

We note that the correspondence between Parseval frames and the Stiefel manifold has appeared previously in the frame theory literature, e.g., [12, 29]. Throughout this section, we equip $\text{St}(n, d)$ with its canonical Riemannian metric (described in the following subsection) and the associated normalized Riemannian volume measure, which we refer to as the *symmetric measure*.

We can now state the main result of this section.

Theorem 3.1 (Concentration over the Stiefel Manifold). *Let $(x_i)_{i=1}^n$ be a random Parseval frame in $\mathbb{R}^d$, drawn from the symmetric measure. Let $\varepsilon > 0$, assume that $n > d \geq 2$, and that n is sufficiently large so that*

$$\frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}} < \varepsilon.$$

Then we have

$$\mathbb{P}\left(\exists i \in \{1, \dots, n\} \text{ such that } \left| \|x_i\| - \sqrt{d/n} \right| \geq \varepsilon\right) \leq \exp\left(-\frac{(n-2)}{8} \left(\varepsilon - \frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}}\right)^2\right).$$

Remark 3.2. *The theorem shows that random Parseval frames concentrate on frames whose norms are all ε -close to $\sqrt{d/n}$, in an additive sense. Frequently in frame theory, an alternative notion of ε -nearly equal-norm frames, based on relative error, is considered. This notion is considered in Section 4, in relation to the Paulsen problem.*

In analogy with Corollary 2.6, the theorem immediately yields a quantification of concentration in n .

Corollary 3.3. *Fix $\varepsilon > 0$. Under the assumptions of Theorem 3.1, suppose further that*

$$\varepsilon \geq \frac{3\sqrt{17}}{2} \sqrt{\frac{d + \ln n}{n}}.$$

If

$$n \geq 2 + \frac{32 \ln(20)}{\varepsilon^2},$$

then the conclusion of Theorem 3.1 holds with confidence at least 95%. If

$$n \geq 2 + \frac{32 \ln(100)}{\varepsilon^2},$$

then the conclusion holds with confidence at least 99%.

The proof of Theorem 3.1 is geometric. It relies on a concentration inequality on manifolds together with two key ingredients: an upper bound for the expectation of a suitable Lipschitz function, and a lower bound on the Ricci curvature of $\text{St}(n, d)$.

Theorem 3.4 ([25]). *Let M be a compact, connected Riemannian manifold with $\text{Ric} \geq \kappa > 0$, and let $f : M \rightarrow \mathbb{R}$ be Lipschitz with constant $\text{Lip}(f)$. If X is distributed according to the normalized Riemannian volume measure on M , then for all $t > 0$,*

$$\mathbb{P}(f(X) - \mathbb{E}[f(X)] \geq t) \leq \exp\left(-\frac{\kappa t^2}{2\text{Lip}(f)^2}\right).$$

The rest of this section is devoted to the proof of Theorem 3.1, which is more technical than the proofs of the concentration results presented in the previous section. We note that a concentration result can be derived somewhat straightforwardly from Theorem 3.4 together with a union bound. However, the proof of Theorem 3.1 avoids the use of a union bound and thereby yields a stronger concentration estimate; see Remark 3.11 for details.

3.1 Geometric Ingredients

Since the main result follows directly from Theorem 3.4, we need two geometric ingredients: the geodesic distance and the Ricci curvature for the Stiefel manifold. For the geodesic distance, it is enough for our purposes to understand its relationship with the Frobenius norm. Indeed, we will define a function whose Lipschitz constant can be estimated easily with respect to the Frobenius norm, and the comparison between the Frobenius norm and the geodesic distance will then allow us to obtain the required Lipschitz constant. For the Ricci curvature, it will be sufficient to establish an appropriate lower bound.

Let us recall some basic geometric structures of the Stiefel manifold; a standard reference for details is [13]. The tangent space at $U \in \text{St}(n, d)$ is given by

$$T_U \text{St}(n, d) = \left\{ \Delta \in \mathbb{R}^{n \times d} \mid U^* \Delta + \Delta^* U = 0 \right\}.$$

Equivalently, any tangent vector $\Delta \in T_U \text{St}(n, d)$ can be expressed in the form

$$\Delta = UA + U_\perp B,$$

where A is a skew-symmetric $d \times d$ matrix, $B \in \mathbb{R}^{(n-d) \times d}$ is an arbitrary matrix, and $U_\perp \in \mathbb{R}^{n \times (n-d)}$ is an orthonormal basis for the complement of the span of the columns of U . The canonical Riemannian metric is defined at $U \in \text{St}(n, d)$ as the map

$$T_U \text{St}(n, d) \times T_U \text{St}(n, d) \rightarrow \mathbb{R}$$

$$(\Delta_1, \Delta_2) \mapsto \text{Tr} \left(\Delta_1^* \left(I_n - \frac{1}{2} U U^* \right) \Delta_2 \right).$$

By general principles, this Riemannian metric yields a volume measure, which we normalize to define the *symmetric measure*, used in the statement of Theorem 3.1. We use d_C to denote the geodesic distance with respect to the canonical metric. As stated above, to deduce that certain functions are Lipschitz with respect to d_C , the following bound in terms of Frobenius distance will suffice.

Lemma 3.5. *For $U_1, U_2 \in \text{St}(n, d)$,*

$$d_C(U_1, U_2) \geq \frac{1}{\sqrt{2}} \|U_1 - U_2\|_F.$$

Proof. By [27],

$$d_C(U_1, U_2) \geq \sqrt{\frac{1}{2}} \cdot 2\sqrt{d} \arcsin \left(\frac{\|U_1 - U_2\|_F}{2\sqrt{d}} \right).$$

The result follows, since $\arcsin(x) \geq x$ for $x \in [0, 1]$. □

Next, we derive a lower bound on Ricci curvature.

Lemma 3.6. *Let $\xi \in T_U \text{St}(n, d)$ be a unit tangent vector with respect to the canonical metric. If $n > d + 1$, then*

$$\text{Ric}(\xi, \xi) \geq \frac{n}{2} - 1.$$

Proof. Write $\xi = UA + U_\perp B$, where A is skew-symmetric and $B \in \mathbb{R}^{(n-d) \times d}$. The explicit formula for the Ricci curvature under the canonical metric, obtained from [31], gives

$$\text{Ric}(\xi, \xi) = \left(\frac{2-d}{4} + \frac{d-n}{4} \right) \text{Tr}(A^2) + \left(\frac{1-d}{2} + (n-2) \right) \text{Tr}(B^* B).$$

Since $\text{Tr}(A^2) = -\|A\|^2$, this becomes

$$\text{Ric}(\xi, \xi) = \frac{n-2}{4} \|A\|^2 + \left(n - \frac{d}{2} - \frac{3}{2} \right) \|B\|^2.$$

Because ξ is a unit vector with respect to the canonical metric,

$$\frac{1}{2} \|A\|^2 + \|B\|^2 = 1.$$

Optimizing under this constraint yields

$$\text{Ric}(\xi, \xi) \geq \min \left\{ \frac{n}{2} - 1, n - \frac{d}{2} - \frac{3}{2} \right\}.$$

The result follows if $n > d + 1$. □

3.2 Analytic Ingredients

We now introduce the function that will be used in the proof of Theorem 3.1. We will estimate its Lipschitz constant and compute its expected value. For $G \in \mathbb{R}^{n \times d}$, define

$$f(G) := \max_{1 \leq i \leq n} \|g_i\|,$$

where $g_i \in \mathbb{R}^d$ denotes the i -th row of G . When restricted to $\text{St}(n, d)$, this function records the maximum row norm of the corresponding Parseval frame.

We first estimate $\mathbb{E}[f(U)]$, where U is drawn from the symmetric measure on $\text{St}(n, d)$. Since the row norms of U are not independent, we compare U to a Gaussian matrix using the following result of Tropp.

Theorem 3.7 ([33, Theorem 1]). *Let U be a random element of $\text{St}(n, d)$ drawn from the symmetric measure, and let $G \in \mathbb{R}^{n \times d}$ be a random matrix with independent $\mathcal{N}(0, n^{-1})$ entries. For each nonnegative, sublinear, convex function f on $\mathbb{R}^{n \times d}$ and each weakly increasing, convex function $\Phi : \mathbb{R} \rightarrow \mathbb{R}$, we have*

$$\mathbb{E} \Phi(f(U)) \leq \mathbb{E} \Phi\left(\left(1 + \frac{d}{2n}\right) f(G)\right).$$

We are now ready to bound the expected value.

Proposition 3.8. *Assume $n > d \geq 2$. Then*

$$\mathbb{E}[f(U)] \leq \frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}}.$$

Proof. Take Φ to be the identity function in Theorem 3.7, which is weakly increasing and convex. The function f is nonnegative, convex, and sublinear on $\mathbb{R}^{n \times d}$, hence

$$\mathbb{E}[f(U)] \leq \left(1 + \frac{d}{2n}\right) \mathbb{E}[f(G)].$$

We now compute the expected value of $f(G)$. This follows from a straightforward calculation, since the vectors are independent and the distribution of each vector can be identified explicitly. $\|g_i\|^2 \sim \frac{1}{n} \chi_d^2$ since each component of g_i has $\mathcal{N}(0, \frac{1}{n})$.

Since $f(G) \geq 0$, we use the standard identity

$$\mathbb{E}[f(G)] = \int_0^\infty \mathbb{P}(f(G) > t) dt.$$

Moreover, by independence,

$$\mathbb{P}(f(G) > t) = 1 - (\mathbb{P}(\|g_i\| \leq t))^n = 1 - \left(\mathbb{P}\left(\frac{1}{n} \chi_d^2 \leq t^2\right)\right)^n = 1 - (\mathbb{P}(\chi_d^2 \leq nt^2))^n.$$

Using $1 - x^n \leq n(1 - x)$ for $x \in [0, 1]$, we obtain

$$\mathbb{P}(f(G) > t) \leq n \mathbb{P}(\chi_d^2 \geq nt^2).$$

Hence, for every $A > 0$,

$$\begin{aligned}\mathbb{E}[f(G)] &= \int_0^A \mathbb{P}(f(G) > t) dt + \int_A^\infty \mathbb{P}(f(G) > t) dt \\ &\leq \int_0^A 1 dt + \int_A^\infty n \mathbb{P}(\chi_d^2 \geq nt^2) dt \\ &= A + n \int_A^\infty \mathbb{P}(\chi_d^2 \geq nt^2) dt.\end{aligned}$$

Now apply Laurent–Massart’s inequality [24, Lemma 1]: if $X \sim \chi_d^2$, then for all $x \geq 0$,

$$\mathbb{P}\left(X - d \geq 2\sqrt{dx} + 2x\right) \leq e^{-x}.$$

Solving $2\sqrt{dx} + 2x = nt^2 - d$ for x yields

$$x = \frac{\left(\sqrt{2nt^2 - d} - \sqrt{d}\right)^2}{4},$$

and therefore

$$\mathbb{P}(\chi_d^2 \geq nt^2) \leq \exp\left(-\frac{\left(\sqrt{2nt^2 - d} - \sqrt{d}\right)^2}{4}\right).$$

Substituting this into the previous inequality and performing the change of variables

$$r = \frac{\sqrt{2nt^2 - d} - \sqrt{d}}{2}$$

gives

$$\mathbb{E}[f(G)] \leq A + \int_B^\infty \frac{2r + \sqrt{d}}{\sqrt{\frac{(2r + \sqrt{d})^2 + d}{2n}}} e^{-r^2} dr, \quad \text{where } A = \sqrt{\frac{(2B + \sqrt{d})^2 + d}{2n}}.$$

Bounding the prefactor by $\sqrt{2n}$, we obtain

$$\mathbb{E}[f(G)] \leq A + \sqrt{2n} \int_B^\infty e^{-r^2} dr.$$

Using

$$\int_B^\infty e^{-r^2} dr \leq \frac{e^{-B^2}}{2B}, \quad B > 0,$$

we conclude that

$$\mathbb{E}[f(G)] \leq \sqrt{\frac{(2B + \sqrt{d})^2 + d}{2n}} + \sqrt{2n} \frac{e^{-B^2}}{2B}.$$

Choosing $B = \sqrt{\ln n}$ yields

$$\mathbb{E}[f(G)] \leq \sqrt{\frac{4 \ln n + 4\sqrt{d \ln n} + 2d}{2n}} + \frac{1}{\sqrt{2n \ln n}}.$$

Since $n > d \geq 2$, we have

$$\begin{aligned}
\mathbb{E}[f(U)] &\leq \left(1 + \frac{d}{2n}\right) \left(\sqrt{\frac{4 \ln n + 4\sqrt{d \ln n} + 2d}{2n}} + \frac{1}{\sqrt{2n \ln n}} \right) \\
&= \frac{1}{\sqrt{n}} \left(1 + \frac{d}{2n}\right) \left(\sqrt{d + 2 \ln n + 2\sqrt{d \ln n}} + \frac{1}{\sqrt{2 \ln n}} \right) \\
&\leq \frac{1}{\sqrt{n}} \left(1 + \frac{d}{2n}\right) \left(\sqrt{d} + \sqrt{2 \ln n} + \frac{1}{\sqrt{2 \ln n}} \right) \\
&\leq \frac{3}{2\sqrt{n}} \left(\frac{3}{2} \sqrt{d} + \sqrt{2 \ln n} \right) \\
&\leq \frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}}.
\end{aligned}$$

This gives the claimed result. $\square$

We next compute a Lipschitz constant for f on $\text{St}(n, d)$.

Proposition 3.9. *The restriction of f to $\text{St}(n, d)$ is $\sqrt{2}$ -Lipschitz with respect to the geodesic distance d_G .*

Proof. Let $U, V \in \text{St}(n, d)$, and let u_i and v_i denote their i -th rows. Then

$$|f(U) - f(V)| = \left| \max_i \|u_i\| - \max_i \|v_i\| \right| \leq \max_i \left| \|u_i\| - \|v_i\| \right| \leq \max_i \|u_i - v_i\|.$$

Therefore,

$$|f(U) - f(V)| \leq \|(\|u_1 - v_1\|, \dots, \|u_n - v_n\|)\|_\infty \leq \|(\|u_1 - v_1\|, \dots, \|u_n - v_n\|)\|_2 = \|U - V\|_F.$$

Thus f is 1-Lipschitz with respect to the Frobenius norm. Applying Lemma 3.5 yields the result. $\square$

3.3 Proof of Theorem 3.1

We are now prepared to prove the main theorem of this section.

Proof of Theorem 3.1. Let $U \in \text{St}(n, d)$ be the matrix associated with the random Parseval frame, and let u_i denote its rows. Since

$$\frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}} < \varepsilon,$$

we have $\varepsilon > \sqrt{\frac{d}{n}}$. Therefore, $\left| \|u_i\| - \sqrt{\frac{d}{n}} \right| \geq \varepsilon$ can only occur when $\|u_i\| - \sqrt{\frac{d}{n}} \geq \varepsilon$, because the alternative inequality $\sqrt{\frac{d}{n}} - \|u_i\| \geq \varepsilon$ would force $\|u_i\| \leq \sqrt{\frac{d}{n}} - \varepsilon < 0$, which is impossible. Hence

$$\mathbb{P}\left(\exists i \text{ such that } \left| \|u_i\| - \sqrt{\frac{d}{n}} \right| \geq \varepsilon\right) = \mathbb{P}\left(\max_{1 \leq i \leq n} \|u_i\| - \sqrt{\frac{d}{n}} \geq \varepsilon\right).$$

Since $f(U) = \max_i \|u_i\|$, this equals

$$\mathbb{P}\left(f(U) - \sqrt{\frac{d}{n}} \geq \varepsilon\right) = \mathbb{P}\left(f(U) - \mathbb{E}[f(U)] \geq \varepsilon - (\mathbb{E}[f(U)] - \sqrt{\frac{d}{n}})\right).$$

By Proposition 3.8,

$$\mathbb{E}[f(U)] \leq \frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}},$$

and therefore

$$\mathbb{P}\left(\exists i \text{ such that } \left|\|u_i\| - \sqrt{\frac{d}{n}}\right| \geq \varepsilon\right) \leq \mathbb{P}\left(f(U) - \mathbb{E}[f(U)] \geq \varepsilon - \frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}}\right).$$

By Proposition 3.9, f is $\sqrt{2}$ -Lipschitz. Moreover, by Lemma 3.6, the Ricci curvature of $\text{St}(n, d)$ under the canonical metric is bounded below by $\frac{n}{2} - 1$. Applying Theorem 3.4 with

$$\kappa = \frac{n}{2} - 1 \quad \text{and} \quad \text{Lip}(f) = \sqrt{2}$$

gives

$$\mathbb{P}(f(U) - \mathbb{E}[f(U)] \geq t) \leq \exp\left(-\frac{(n-2)t^2}{8}\right), \quad t \geq 0.$$

Finally, substituting

$$t = \varepsilon - \frac{3\sqrt{17}}{4} \sqrt{\frac{d + \ln n}{n}},$$

which is positive by assumption, yields the claim. $\square$

In the proof above, we employed an explicit upper bound on $\mathbb{E}[f(G)]$. In principle, one could obtain a sharper estimate by using the exact expression for $\mathbb{E}[f(G)]$, which involves the incomplete gamma function and the gamma function. Since our goal was to state a clean explicit concentration bound and keep the resulting inequality as simple as possible, we used an upper bound for $\mathbb{E}[f(G)]$. For completeness, we record below the concentration result obtained when the exact expression is used.

Proposition 3.10. *Let $(x_i)_{i=1}^n$ be a random Parseval frame in $\mathbb{R}^d$, drawn from the symmetric measure, and let $\varepsilon > 0$ be given. Then, for all sufficiently large n satisfying $K < \varepsilon$, we have*

$$\mathbb{P}\left(\max_{i \in \{1, \dots, n\}} \|x_i\| - \sqrt{\frac{d}{n}} \geq \varepsilon\right) \leq \exp\left(-\frac{(n-2)(\varepsilon - K)^2}{8}\right),$$

where

$$K := \left(1 + \frac{d}{2n}\right) \frac{1}{\sqrt{2n}} \int_0^\infty \frac{1 - \left(\frac{\gamma(\frac{d}{2}, y)}{\Gamma(\frac{d}{2})}\right)^n}{\sqrt{y}} dy - \sqrt{\frac{d}{n}}.$$

Although this formula provides a sharper result, we chose to state the previous version instead. The reason is that the constant K appearing here is rather complicated, whereas in the previous formulation we replace it by a simpler upper bound that can be stated more transparently. If $n \geq d/\varepsilon^2$, the one-sided condition appearing in the probability estimate in Proposition 3.10 is equivalent to the corresponding two-sided condition. In Section A, we compare the theoretical predictions of Proposition 3.10 with empirical results obtained from numerical experiments.

Remark 3.11. *An alternative concentration estimate can be obtained by applying Theorem 3.4 to each coordinate separately and then using a union bound. Define*

$$g_i : \text{St}(n, d) \rightarrow \mathbb{R}, \quad g_i(U) = e_i^* \text{diag}(UU^*),$$

so that $g_i(U) = \|u_i\|^2$. Then $\mathbb{E}[g_i(U)] = d/n$, and g_i is $2\sqrt{2}$ -Lipschitz with respect to the geodesic distance. Applying Theorem 3.4 and summing over i yields

$$\mathbb{P}\left(\exists i, \left|\|u_i\|^2 - \frac{d}{n}\right| \geq \varepsilon\right) \leq 2n \exp\left(-\frac{(n-2)\varepsilon^2}{32}\right).$$

This bound is valid, but weaker than Theorem 3.1 because of the extra factor n produced by the union bound.

4 Application to the Paulsen Problem

In this section, we present an application of Theorem 2.1 to the *Paulsen problem*, a major open problem in finite-dimensional frame theory [3, 4, 7, 18, 22], which we now recall. We call $(x_i)_{i=1}^n$ an ε -nearly Parseval frame if $\|S - I_d\|_{\text{op}} \leq \varepsilon$, and an ε -nearly equal-norm frame if the squared norms of its frame vectors satisfy $(1 - \varepsilon)\frac{d}{n} \leq \|x_i\|^2 \leq (1 + \varepsilon)\frac{d}{n}$ for all $i \in \{1, \dots, n\}$. We write $\mathcal{F}_{d,n}^\varepsilon$ for the class of frames that are both ε -nearly equal-norm and ε -nearly Parseval, and $\mathcal{F}_{d,n}$ for the class of (exactly) equal-norm Parseval frames in $\mathbb{R}^d$.

Problem 4.1 (Paulsen Problem [3]). *Find the simplest function $f(d, n, \varepsilon)$ satisfying, for all $(x_i)_{i=1}^n \in \mathcal{F}_{d,n}^\varepsilon$, the bound*

$$\inf_{(y_i)_{i=1}^n \in \mathcal{F}_{d,n}} d((x_i)_{i=1}^n, (y_i)_{i=1}^n)^2 \leq f(d, n, \varepsilon),$$

where the distance is

$$d((x_i)_{i=1}^n, (y_i)_{i=1}^n) = \left(\sum_{i=1}^n \|x_i - y_i\|^2\right)^{1/2}.$$

The existence of such a function $f(d, n, \varepsilon)$ is attributed to Hadwin by Bodmann and Casazza [3]. Subsequent work focused on expressing a bounding function in the cleanest possible form and, in particular, without explicit dependence on n . In [22], Kwok et al. were the first to obtain a bound depending only on d and ε ; more precisely, they proved a bound of order $O(\varepsilon d^{13/2})$. Subsequently, Hamilton and Moitra obtained the following result, which will be useful for our purposes.

Theorem 4.2 ([18]). *For $(x_i)_{i=1}^n \in \mathcal{F}_{d,n}^\varepsilon$,*

$$\inf_{(y_i)_{i=1}^n \in \mathcal{F}_{d,n}} d((x_i)_{i=1}^n, (y_i)_{i=1}^n)^2 \leq 20\varepsilon d^2.$$

At present, this is the best general bound for the Paulsen problem in the literature. On the other hand, it was shown in [7] that any such bound must necessarily be at least of order $\varepsilon^2 d$. Our main result is the following. In the statement, we refer to a *uniform measure* on $\mathcal{F}_{d,n}^\varepsilon$, which is defined precisely in the following subsection—see (3).

Theorem 4.3 (Concentration for the Paulsen Problem). *Let $\varepsilon \in (0, 1)$ and let $(x_i)_{i=1}^n$ be chosen uniformly at random from $\mathcal{F}_{d,n}^\varepsilon$. Then the bound*

$$\inf_{(y_i)_{i=1}^n \in \mathcal{F}_{d,n}} d((x_i)_{i=1}^n, (y_i)_{i=1}^n)^2 \leq (\sqrt{20} + \sqrt{2})^2 \varepsilon^2 d \quad (2)$$

holds with probability at least

$$1 - \frac{2d \exp\left(-\frac{n\varepsilon^4}{2d^3}\right)}{1 - 2d \exp\left(-\frac{6\varepsilon^2 n}{5(d+2)^2}\right)}.$$

The next (immediate) corollary quantifies the success probability in terms of n .

Corollary 4.4. *A random frame in $\mathcal{F}_{d,n}^\varepsilon$ satisfies the Paulsen bound (2) with probability 95% if*

$$n \geq \frac{2d^3 \ln(80d)}{\varepsilon^4},$$

and with probability 99% if

$$n \geq \frac{2d^3 \ln(400d)}{\varepsilon^4}.$$

The theorem is proven in the next subsection, after setting up necessary preliminary results. In Section 4.2, we clarify the distinction between the best possible deterministic scaling and the high-probability scaling obtained here.

4.1 Proof of Theorem 4.3

Let us begin by setting up some relevant notation. Consider the subset of $\mathbb{R}^d$ consisting of all vectors whose squared norms lie between $(1 - \varepsilon)\frac{d}{n}$ and $(1 + \varepsilon)\frac{d}{n}$. This set can be viewed as a thin spherical shell in $\mathbb{R}^d$. We denote it by

$$\mathcal{S}_\varepsilon^d := \left\{ v \in \mathbb{R}^d : (1 - \varepsilon)\frac{d}{n} \leq \|v\|^2 \leq (1 + \varepsilon)\frac{d}{n} \right\}.$$

Let μ denote the uniform probability measure on $\mathcal{S}_\varepsilon^d$, and consider the product measure $\mu^{\otimes n}$ on $(\mathcal{S}_\varepsilon^d)^n$. Our primary set of interest is $\mathcal{F}_{d,n}^\varepsilon \subset (\mathcal{S}_\varepsilon^d)^n$. Since $\mathbb{P}_{\mu^{\otimes n}}((x_1, \dots, x_n) \in \mathcal{F}_{d,n}^\varepsilon) > 0$, the corresponding conditional probabilities are well-defined. Let ν denote the conditional probability measure, i.e.,

$$\nu(\cdot) := \mu^{\otimes n}(\cdot \mid \mathcal{F}_{d,n}^\varepsilon). \quad (3)$$

Moreover, ν is precisely the normalized restriction of the uniform product measure to $\mathcal{F}_{d,n}^\varepsilon$; in particular, it is again uniform on $\mathcal{F}_{d,n}^\varepsilon$. This is the measure on $\mathcal{F}_{d,n}^\varepsilon$ that is used in Theorem 4.3.

The proof of Theorem 4.3 consists of two separate parts. In the first part, we compute the relevant probability, and in the second part, we derive the upper bound. Since the proof is rather lengthy, we first establish some auxiliary results that will be needed in the argument.

Lemma 4.5. *Let v be a random vector uniformly distributed on $\mathcal{S}_\varepsilon^d$. Then,*

$$\mathbb{E}\|v\|^2 = \frac{d}{n} \cdot \frac{d}{d+2} \cdot \frac{(1+\varepsilon)^{\frac{d+2}{2}} - (1-\varepsilon)^{\frac{d+2}{2}}}{(1+\varepsilon)^{\frac{d}{2}} - (1-\varepsilon)^{\frac{d}{2}}} \leq \frac{d}{n} \cdot \frac{d}{d+2} \cdot \left(1 + \varepsilon + \frac{2(1-\varepsilon)}{d}\right).$$

Proof. By the definition of the expected value, we have

$$\mathbb{E}\|v\|^2 = \frac{\int_{S_\varepsilon^d} \|v\|^2 dv}{\int_{S_\varepsilon^d} dv}.$$

Using polar coordinates $v = r\omega$ with $\omega \in S^{d-1}$ and $dv = r^{d-1} dr d\omega$, and writing

$$a = \sqrt{\frac{d}{n}(1-\varepsilon)}, \quad b = \sqrt{\frac{d}{n}(1+\varepsilon)},$$

we obtain

$$\mathbb{E}\|v\|^2 = \frac{\int_a^b r^2 r^{d-1} dr}{\int_a^b r^{d-1} dr} = \frac{d}{d+2} \cdot \frac{b^{d+2} - a^{d+2}}{b^d - a^d}.$$

Substituting a, b yields

$$\mathbb{E}\|v\|^2 = \frac{d}{n} \cdot \frac{d}{d+2} \cdot \frac{(1+\varepsilon)^{\frac{d+2}{2}} - (1-\varepsilon)^{\frac{d+2}{2}}}{(1+\varepsilon)^{\frac{d}{2}} - (1-\varepsilon)^{\frac{d}{2}}}, \quad (4)$$

and this proves the equality part of the lemma.

Since the expression (4) is complicated, we now derive a simpler upper bound. Let $k = d/2$, $\alpha = 1 + \varepsilon$ and $\beta = 1 - \varepsilon$ where $\varepsilon \in (0, 1)$. Define

$$L := \frac{\alpha^{k+1} - \beta^{k+1}}{\alpha^k - \beta^k}.$$

Observe that $\alpha^{k+1} - \beta^{k+1} = \alpha(\alpha^k - \beta^k) + \beta^k(\alpha - \beta)$, hence

$$L = \alpha + \frac{\beta^k(\alpha - \beta)}{\alpha^k - \beta^k} = \alpha + \frac{\alpha - \beta}{\left(\frac{\alpha}{\beta}\right)^k - 1}.$$

Now

$$\alpha - \beta = 2\varepsilon, \quad \frac{\alpha}{\beta} = \frac{1+\varepsilon}{1-\varepsilon} = 1 + u, \quad \text{where } u = \frac{2\varepsilon}{1-\varepsilon} > 0.$$

Then

$$\left(\frac{\alpha}{\beta}\right)^k - 1 = (1+u)^k - 1 \geq ku \quad \text{by Bernoulli's inequality.}$$

Therefore,

$$\frac{\alpha - \beta}{\left(\frac{\alpha}{\beta}\right)^k - 1} \leq \frac{2\varepsilon}{k \cdot \frac{2\varepsilon}{1-\varepsilon}} = \frac{1-\varepsilon}{k}.$$

Thus

$$L \leq \alpha + \frac{1-\varepsilon}{k}.$$

This implies the inequality in the lemma. □

Using Lemma 4.5, we obtain the following result.

Proposition 4.6. *Let $x_1, \dots, x_n$ be i.i.d. sampled from $\mathcal{S}_\varepsilon^d$. Then*

$$\mathbb{P}_{\mu^{\otimes n}}((x_1, \dots, x_n) \in \mathcal{F}_{d,n}^\varepsilon) > 1 - 2d \exp\left(-\frac{24\varepsilon^2 n}{(d+2)(1+\varepsilon)(3(d+2)(1+\varepsilon) + 4d\varepsilon)}\right).$$

Proof. We have

$$\begin{aligned} \mathbb{P}_{\mu^{\otimes n}}((x_1, \dots, x_n) \in \mathcal{F}_{d,n}^\varepsilon) &= \mathbb{P}_{\mu^{\otimes n}}((x_1, \dots, x_n) \text{ is } \varepsilon\text{-nearly Parseval}) \\ &= \mathbb{P}_{\mu^{\otimes n}}(\|S - I_d\|_{\text{op}} \leq \varepsilon) \\ &= 1 - \mathbb{P}_{\mu^{\otimes n}}(\|S - I_d\|_{\text{op}} > \varepsilon) \end{aligned}$$

Proceeding as in the sphere and ball cases, we invoke Theorem 2.2, the Matrix Bernstein inequality. Define $Y_i := x_i x_i^* - \mathbb{E}[x_i x_i^*]$, so that, clearly $\mathbb{E}[Y_i] = 0$, and $\mathbb{E}[x_i x_i^*] = (\mathbb{E}\|v\|^2/d)I_d$ by the rotation invariant distribution. Next, we estimate the parameters in Theorem 2.2.

First, we need to find R , an upper bound for $\|Y_i\|_{\text{op}}$. Since $\|x_i\|^2 \leq \frac{d}{n}(1+\varepsilon)$ for all $x_i \in \mathcal{S}_\varepsilon^d$, we have

$$\|x_i x_i^*\|_{\text{op}} = \|x_i\|^2 \leq \frac{d}{n}(1+\varepsilon).$$

Therefore,

$$\|Y_i\|_{\text{op}} = \left\| x_i x_i^* - \left(\frac{\mathbb{E}\|v\|^2}{d} \right) I_d \right\|_{\text{op}} = \frac{d}{n}(1+\varepsilon) - \frac{\mathbb{E}\|v\|^2}{d} \leq \frac{d}{n}(1+\varepsilon),$$

so we can take the last term as R .

Now we derive an expression for the variance parameter σ^2 . We compute

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[Y_i^2] &= n \mathbb{E}[Y_i^2] = n \mathbb{E} \left[\|x_i\|^2 x_i x_i^* - 2 \left(\frac{\mathbb{E}\|x_i\|^2}{d} \right) x_i x_i^* + \left(\frac{\mathbb{E}\|x_i\|^2}{d} \right)^2 I_d \right] \\ &= n \left(\frac{\mathbb{E}[\|x_i\|^4]}{d} - \frac{(\mathbb{E}\|x_i\|^2)^2}{d^2} \right) I_d. \end{aligned}$$

Taking operator norms and using $\|x_i\|^2 \leq \frac{d}{n}(1+\varepsilon)$, we obtain

$$\sigma^2 := \left\| \sum_{i=1}^n \mathbb{E}[Y_i^2] \right\|_{\text{op}} \leq n \left\| \frac{\mathbb{E}[\|x_i\|^4]}{d} I_d \right\|_{\text{op}} \leq n \cdot \frac{\left(\frac{d}{n}(1+\varepsilon) \right)^2}{d} = \frac{d}{n}(1+\varepsilon)^2.$$

Therefore, we obtain

$$\mathbb{P}(\|S - M I_d\|_{\text{op}} > t) \leq 2d \exp\left(-\frac{3t^2 n}{6d(1+\varepsilon)^2 + 2d(1+\varepsilon)t}\right),$$

where $M := n(\mathbb{E}\|v\|^2/d)$. By the triangle inequality for the operator norm, if

$$\|S - I_d\|_{\text{op}} - \|I_d - M I_d\|_{\text{op}} > t,$$

then

$$\|S - MI_d\|_{\text{op}} > t.$$

Hence, the probability of the first event is bounded by the probability of the second, and the previous tail bound also applies to

$$\mathbb{P}\left(\|S - I_d\|_{\text{op}} > t + \|I_d - MI_d\|_{\text{op}}\right).$$

Since our aim is an ε -nearly Parseval statement, we set

$$\varepsilon = t + \|I_d - MI_d\|_{\text{op}}, \quad \text{i.e.,} \quad t = \varepsilon - \|I_d - MI_d\|_{\text{op}}.$$

To use the bound we must have $t > 0$, so it remains to ensure

$$\varepsilon - \|I_d - MI_d\|_{\text{op}} = \varepsilon - |1 - M| > 0.$$

By Lemma 4.5, we know the exact value and an upper bound for M

$$M = \frac{d}{d+2} \cdot \frac{(1+\varepsilon)^{\frac{d+2}{2}} - (1-\varepsilon)^{\frac{d+2}{2}}}{(1+\varepsilon)^{\frac{d}{2}} - (1-\varepsilon)^{\frac{d}{2}}} \leq \frac{d}{d+2} \cdot \left(1 + \varepsilon + \frac{2(1-\varepsilon)}{d}\right).$$

Viewing M as a function of d , one checks that it is increasing in d and satisfies $M = 1$ at $d = 2$. Hence, for $d \geq 2$ we have $M \geq 1$, and therefore

$$\|I_d - MI_d\|_{\text{op}} = |1 - M| = M - 1.$$

In particular,

$$t = \varepsilon - \|I_d - MI_d\|_{\text{op}} = \varepsilon - (M - 1) = \varepsilon + 1 - M \geq \frac{4\varepsilon}{d+2} > 0.$$

Substituting all of the parameters we have obtained into Proposition 2.2 yields the claimed result. □

We now have all the necessary ingredients to prove Theorem 4.3.

Proof of Theorem 4.3. We define the radial projection map

$$P : \mathcal{S}_\varepsilon^d \longrightarrow \sqrt{\frac{d}{n}} \mathbb{S}^{d-1} \quad \text{by} \quad P(v) = \frac{\sqrt{d/n}}{\|v\|} v.$$

This map rescales each vector in the shell onto the sphere of radius $\sqrt{d/n}$. Since μ is rotation-invariant and $P(Qv) = QP(v)$ for every orthogonal transformation $Q \in O(d)$, it follows that the pushforward measure $P_\# \mu$ is also rotation-invariant on the sphere $\sqrt{d/n} \mathbb{S}^{d-1}$. By uniqueness of the normalized Haar measure on the sphere, we conclude that $P_\# \mu$ is precisely the uniform probability measure on $\sqrt{d/n} \mathbb{S}^{d-1}$.

The strategy for the rest of the proof is sketched as follows. After selecting $(x_i)_{i=1}^n \in \mathcal{F}_{d,n}^\varepsilon$, we show that the projected vectors $(P(x_i))_{i=1}^n$ are, with high probability, ε^2/d -nearly Parseval. Since they also have equal-norm with $\|P(x_i)\| = \sqrt{d/n}$, they belong to $\mathcal{F}_{d,n}^{\varepsilon^2/d}$, so that we can apply the result of Hamilton and Moitra (Theorem 4.2), we obtain a sharper upper bound. Moreover, we show that $(x_i)_{i=1}^n$ is close to $(P(x_i))_{i=1}^n$. We will then apply the triangle inequality to conclude the proof.

We begin by determining the relevant probability. For this purpose, we introduce some notation and definitions.

$$\mathcal{A} := \left\{ \frac{\varepsilon^2}{d}\text{-nearly Parseval frames with vectors in } \sqrt{\frac{d}{n}} \mathbb{S}^{d-1} \right\}.$$

We are interested in the conditional probability $\mathbb{P}_\nu \left((P(x_1), \dots, P(x_n)) \in \mathcal{A} \right)$. Since $\mathcal{F}_{d,n}^\varepsilon$ corresponds exactly to the ε -nearly Parseval condition in $(\mathcal{S}_\varepsilon^d)^n$, this equals

$$\mathbb{P}_{\mu^{\otimes n}} \left((P(x_1), \dots, P(x_n)) \in \mathcal{A} \mid (x_1, \dots, x_n) \text{ is } \varepsilon\text{-nearly Parseval} \right).$$

Using complements, we write

$$= 1 - \mathbb{P}_{\mu^{\otimes n}} \left((P(x_1), \dots, P(x_n)) \notin \mathcal{A} \mid (x_1, \dots, x_n) \text{ is } \varepsilon\text{-nearly Parseval} \right).$$

By the definition of conditional probability,

$$= 1 - \frac{\mathbb{P}_{\mu^{\otimes n}} \left((P(x_1), \dots, P(x_n)) \notin \mathcal{A} \text{ and } (x_1, \dots, x_n) \text{ are } \varepsilon\text{-nearly Parseval} \right)}{\mathbb{P}_{\mu^{\otimes n}} \left((x_1, \dots, x_n) \text{ is } \varepsilon\text{-nearly Parseval} \right)}.$$

Finally, we obtain the lower bound

$$\geq 1 - \frac{\mathbb{P}_{\mu^{\otimes n}} \left((P(x_1), \dots, P(x_n)) \notin \mathcal{A} \right)}{\mathbb{P}_{\mu^{\otimes n}} \left((x_1, \dots, x_n) \text{ is } \varepsilon\text{-nearly Parseval} \right)}.$$

Since $P_\# \mu$ is the uniform probability measure on $\mathbb{S}_{\sqrt{d/n}}^{d-1}$, the vectors $P(x_1), \dots, P(x_n)$ are i.i.d. uniform on the sphere. Hence

$$\mathbb{P}_{\mu^{\otimes n}} \left((P(x_1), \dots, P(x_n)) \notin \mathcal{A} \right)$$

is exactly the probability in Theorem 2.1 with ε replaced by ε^2/d .

Therefore,

$$\mathbb{P}_{\mu^{\otimes n}} \left((P(x_1), \dots, P(x_n)) \notin \mathcal{A} \right) \leq 2d \exp \left(- \frac{n\varepsilon^4}{2d \left(d^2 - d + \frac{2\varepsilon^2(d-1)}{3} \right)} \right).$$

By Proposition 4.6, we have

$$\mathbb{P}_{\mu^{\otimes n}}\left((x_1, \dots, x_n) \text{ is } \varepsilon\text{-nearly Parseval}\right) > 1 - 2d \exp\left(-\frac{24\varepsilon^2 n}{(d+2)(1+\varepsilon)(3(d+2)(1+\varepsilon)+4d\varepsilon)}\right)$$

Consequently, we obtain the following lower bound for the main probability:

$$\mathbb{P}_\nu\left((P(x_1), \dots, P(x_n)) \in \mathcal{A}\right) \geq 1 - \frac{2d \exp\left(-\frac{n\varepsilon^4}{2d\left(d^2-d+\frac{2\varepsilon^2(d-1)}{3}\right)}\right)}{1 - 2d \exp\left(-\frac{24\varepsilon^2 n}{(d+2)(1+\varepsilon)(3(d+2)(1+\varepsilon)+4d\varepsilon)}\right)}$$

After simplifying the expression in this probability bound, we may write the following alternative lower bound.

$$\mathbb{P}_\nu\left((P(x_1), \dots, P(x_n)) \in \mathcal{A}\right) \geq 1 - \frac{2d \exp\left(-\frac{n\varepsilon^4}{2d^3}\right)}{1 - 2d \exp\left(-\frac{6\varepsilon^2 n}{5(d+2)^2}\right)}$$

We have now completed the analysis of the probability term appearing in the theorem. In particular, for a randomly chosen $(x_i)_{i=1}^n \in \mathcal{F}_{d,n}^\varepsilon$, we have obtained a lower bound on the probability that $(z_i)_{i=1}^n$ with $z_i = Px_i$ forms an ε^2/d -nearly Parseval frame. Moreover, since each z_i satisfies $\|z_i\| = \sqrt{d/n}$, the vectors automatically have equal-norm. Consequently, $(z_i)_{i=1}^n$ belongs to $\mathcal{F}_{d,n}^{\varepsilon^2/d}$ with at least the same probability. In this setting, applying Theorem 4.2 of Hamilton and Moitra yields

$$\inf_{(y_i)_{i=1}^n \in \mathcal{F}_{d,n}} d((y_i)_{i=1}^n, (z_i)_{i=1}^n)^2 \leq 20 \frac{\varepsilon^2}{d} d^2 = 20\varepsilon^2 d.$$

At this point, our main focus is to estimate $d((x_i)_{i=1}^n, (z_i)_{i=1}^n)^2$. We have

$$x_i - z_i = x_i \left(1 - \sqrt{\frac{d}{n}} \frac{1}{\|x_i\|}\right),$$

so that

$$\|x_i - z_i\|^2 = \|x_i\|^2 \left(1 - \sqrt{\frac{d}{n}} \frac{1}{\|x_i\|}\right)^2 = \left(\|x_i\| - \sqrt{\frac{d}{n}}\right)^2.$$

Summing over i , we have

$$\sum_{i=1}^n \|x_i - z_i\|^2 = \sum_{i=1}^n \left(\|x_i\| - \sqrt{\frac{d}{n}}\right)^2.$$

Note that

$$\|x_i\| \in \left[\sqrt{\frac{d}{n}}\sqrt{1-\varepsilon}, \sqrt{\frac{d}{n}}\sqrt{1+\varepsilon}\right],$$

which implies

$$\left|\|x_i\| - \sqrt{\frac{d}{n}}\right| \leq \sqrt{\frac{d}{n}} (\sqrt{1+\varepsilon} - \sqrt{1-\varepsilon}),$$

yielding

$$\sum_{i=1}^n \|x_i - z_i\|^2 \leq \sum_{i=1}^n \frac{d}{n} (\sqrt{1+\varepsilon} - \sqrt{1-\varepsilon})^2 = d (\sqrt{1+\varepsilon} - \sqrt{1-\varepsilon})^2.$$

Next, observe that

$$(\sqrt{1+\varepsilon} - \sqrt{1-\varepsilon})^2 = 2 - 2\sqrt{1-\varepsilon^2} < 2 - 2(1-\varepsilon^2) = 2\varepsilon^2,$$

which implies

$$d((x_i)_{i=1}^n, (z_i)_{i=1}^n)^2 = \sum_{i=1}^n \|x_i - z_i\|^2 \leq 2\varepsilon^2 d.$$

By using triangle inequality, we have

$$\begin{aligned} \inf_{(y_i)_{i=1}^n \in \mathcal{F}_{d,n}} d((x_i)_{i=1}^n, (y_i)_{i=1}^n) &\leq d((x_i)_{i=1}^n, (z_i)_{i=1}^n) + \inf_{(y_i)_{i=1}^n \in \mathcal{F}_{d,n}} d((y_i)_{i=1}^n, (z_i)_{i=1}^n) \\ &\leq \sqrt{2\varepsilon^2 d} + \sqrt{20\varepsilon^2 d} = (\sqrt{2} + \sqrt{20}) \sqrt{\varepsilon^2 d}. \end{aligned}$$

After squaring both sides, we obtain the result. $\square$

It was shown by Cahill and Casazza [4] that the Paulsen problem is equivalent to the *projection problem* in operator theory. Using this equivalence, we have the following probabilistic result for the projection problem. The proof follows directly from [4, Theorem 4.1] and Theorem 4.3.

Corollary 4.7. *Let $e_1, \dots, e_n \in \mathbb{R}^n$ be the standard basis vectors. Let P be a random projection to a d -dimensional subspace, sampled under the condition that*

$$(1-\varepsilon)\frac{d}{n} \leq \|Pe_i\|^2 \leq (1+\varepsilon)\frac{d}{n},$$

for all i , where $\|\cdot\|$ is the standard norm. Then there exists a projection Q with $\|Qe_i\|^2 = \frac{d}{n}$, for all i , satisfying

$$\sum_{i=1}^n \|Pe_i - Qe_i\|^2 \leq 4(\sqrt{20} + \sqrt{2})^2 \varepsilon^2 d$$

with high probability, provided n is sufficiently large.

4.2 On the Optimal Paulsen Bound

The goal of this subsection is to clarify the theoretical order of an optimal bound $f(\varepsilon, d, n)$ in the Paulsen problem. As was mentioned above, it was shown in [7], via a family of examples, that the Paulsen problem must at least be on the order of $\varepsilon^2 d$. On the other hand, quoting from [4]: “For all examples we know at this time, we have $f(\varepsilon, d, n) \leq 16\varepsilon d$.” The papers [22] and [18] state that the optimal theoretical bound in the Paulsen problem must be at least of order εd , both citing [4] as the source. However, as far as we can tell, there is no such theoretical bound established in that paper, or anywhere else in the literature. Our goal is to clarify the situation by proving:

Proposition 4.8. *If a bound $f(\varepsilon, d, n)$ for the Paulsen problem is a polynomial in ε and d , then it must be of order at least $O(\varepsilon d)$.*

Remark 4.9. Theorem 4.3 gives a bound of the form $O(\varepsilon^2 d)$. Since the theorem is probabilistic, this does not contradict the proposition.

The proof proceeds via a family of examples based on the example appearing in Equation (38) of [8], which was later presented in greater detail, together with a proof, in Ramachandran's thesis [32].

Example 4.10 ([32, Example A.1.1]). For a fixed θ which is small enough, consider the synthesis operator U where the columns represent a frame:

$$U := \frac{\sqrt{2}}{2} \begin{bmatrix} \cos(2\theta) & \cos(2\theta) & 0 & 0 \\ \sin(2\theta) & -\sin(2\theta) & 1 & 1 \end{bmatrix}.$$

The corresponding frame of U is in $\mathcal{F}_{2,4}^\varepsilon$ where $\varepsilon \leq 4 \sin^2 \theta$. The closest equal-norm Parseval frame (ENPF) to U has synthesis operator

$$V := \frac{\sqrt{2}}{2} \begin{bmatrix} \cos(\theta) & \cos(-\theta) & \sin(-\theta) & \sin(\theta) \\ \sin(\theta) & \sin(-\theta) & \cos(-\theta) & \cos(\theta) \end{bmatrix},$$

and the squared distance is $\|U - V\|_F^2 \geq \sin^2 \theta \geq \varepsilon/4$.

Remark 4.11. This example is exactly equal norm, and its frame operator is $I + O(\theta^2)$ for small values of θ . Geometrically, this is saying that his family of examples is contained in the equal norm locus and is tangent to the Parseval locus. Thus this is a particularly structured example and in retrospect it should not be surprising that it manages to stay closer to the equal norm and Parseval locus than a generic perturbation.

By using this example, we generate the next one.

Example 4.12. Let k be a positive integer and let $d = 2k$ and $n = 4k$. Let U_k and V_k be the following matrices

$$U_k := \begin{bmatrix} U & 0 & \cdots & 0 \\ 0 & U & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & U \end{bmatrix} \quad \text{and} \quad V_k := \begin{bmatrix} V & 0 & \cdots & 0 \\ 0 & V & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V \end{bmatrix}$$

where U and V are the matrices defined in the previous example. Let ε be the same number from the previous example.

Lemma 4.13. The frames represented by U_k and V_k belong to $\mathcal{F}_{d,n}^\varepsilon$ and $\mathcal{F}_{d,n}$, respectively.

Proof. The frame vectors represented by U_k have the same nonzero components as those of U , with additional zero entries. In particular, each column of U_k has the same norm as the corresponding column of U , so the equal-norm condition is preserved. Moreover, $U_k U_k^*$ is a block diagonal matrix with each diagonal block equal to $U U^*$, and therefore has the same eigenvalues as $U U^*$, each with multiplicity k . Since U represents an ε -nearly equal-norm Parseval frame, all eigenvalues of $U U^*$ lie between $1 - \varepsilon$ and $1 + \varepsilon$, and the same holds for $U_k U_k^*$. Hence U_k represents a frame in $\mathcal{F}_{d,n}^\varepsilon$. The same argument applied to V and V_k shows that V_k represents a frame in $\mathcal{F}_{d,n}$. $\square$

Lemma 4.14. *Let $a_1, a_2 \in \mathbb{R}$ with $a_1^2 + a_2^2 = 1/2$, and let $b_1, b_2, \dots, b_d \in \mathbb{R}$ with $b_1^2 + b_2^2 + \dots + b_d^2 = 1/2$. Then the expression*

$$Y = (a_1 - b_1)^2 + (a_2 - b_2)^2 + \sum_{i=3}^d b_i^2$$

is minimized only when $b_3 = \dots = b_d = 0$.

Proof. This result follows from the Cauchy–Schwarz inequality. $\square$

Lemma 4.15. *The frame represented by V_k is the closest ENPF to the frame represented by U_k .*

Proof. Let H be the synthesis operator of an arbitrary frame in $\mathcal{F}_{2k,4k}$. We decompose H as $H = H^{\text{diag}} + H^{\text{off}}$, where H^{diag} retains the 2×4 diagonal blocks of H and sets all other entries to zero, and H^{off} retains the remaining entries and sets the diagonal blocks to zero. Since U_k is block diagonal, $U_k - H^{\text{diag}}$ and H^{off} have disjoint support, and therefore

$$\|U_k - H\|_{\text{F}}^2 = \|U_k - H^{\text{diag}}\|_{\text{F}}^2 + \|H^{\text{off}}\|_{\text{F}}^2.$$

Denote the columns of U_k by $f_1, \dots, f_{4k}$ and the columns of H by $h_1, \dots, h_{4k}$. For each i , write $h_i = h'_i + h''_i$, where h'_i is the part supported on the diagonal block containing f_i and h''_i is the remaining part. Since $\|f_i\|^2 = \|h_i\|^2 = \frac{1}{2}$ and $\|h_i\|^2 = \|h'_i\|^2 + \|h''_i\|^2$, we have

$$\|U_k - H\|_{\text{F}}^2 = \sum_{i=1}^{4k} (\|f_i - h'_i\|^2 + \|h''_i\|^2).$$

Each summand has the form treated in Lemma 4.14: the nonzero components of f_i play the role of (a_1, a_2) and the components of h_i play the role of $(b_1, \dots, b_d)$, all with fixed squared norm $\frac{1}{2}$. By the lemma, each summand is minimized only when $h''_i = 0$. Therefore the minimizer H must be block diagonal.

Since H is block diagonal and represents an equal-norm Parseval frame, each 2×4 diagonal block H_i of H itself represents an ENPF. Therefore

$$\|U_k - H\|_{\text{F}}^2 = \sum_{i=1}^k \|H_i - U\|_{\text{F}}^2 \geq k \|V - U\|_{\text{F}}^2,$$

where the inequality follows from the fact that the frame represented by V is the closest ENPF to the frame represented by U . This lower bound is attained when $H = V_k$, so the frame represented by V_k is the closest ENPF to the frame represented by U_k . $\square$

Combining the above lemmas yields:

Proposition 4.16. *The minimum squared distance from the frame represented by U_k to $\mathcal{F}_{d,n}$ satisfies*

$$\|V_k - U_k\|_{\text{F}}^2 = k \|V - U\|_{\text{F}}^2 \geq k \cdot \frac{\varepsilon}{4} = \frac{d\varepsilon}{8}.$$

In particular, for any ε -nearly ENPF in the regime $d = 2k$, $n = 4k$, the distance to the closest ENPF grows linearly in ε , showing that the bound of order $\varepsilon^2 d$ cannot hold uniformly over all ε -nearly ENPFs.

This family of examples therefore proves Proposition 4.8.

References

- [1] John J. Benedetto and Matthew Fickus. Finite normalized tight frames. *Advances in Computational Mathematics*, 18:357–385, 2003.
- [2] Bernhard G. Bodmann. Random fusion frames are nearly equiangular and tight. *Linear Algebra and its Applications*, 439(5):1401–1414, 2013.
- [3] Bernhard G. Bodmann and Peter G. Casazza. The road to equal-norm Parseval frames. *Journal of Functional Analysis*, 258(2):397–420, 2010.
- [4] Jameson Cahill and Peter G. Casazza. The Paulsen problem in operator theory. *Operators and Matrices*, 7(1):117–130, 2013.
- [5] Jameson Cahill, Dustin G. Mixon, and Nate Strawn. Connectivity and irreducibility of algebraic varieties of finite unit norm tight frames. *SIAM Journal on Applied Algebra and Geometry*, 1(1):38–72, 2017.
- [6] Peter G. Casazza. The art of frame theory. *Taiwanese Journal of Mathematics*, 4(2):129–201, 2000.
- [7] Peter G. Casazza. The Kadison–Singer and Paulsen problems in finite frame theory. In *Finite Frames: Theory and Applications*, Applied and Numerical Harmonic Analysis, pages 381–413. Birkhäuser/Springer, New York, 2013.
- [8] Peter G. Casazza, Matthew Fickus, and Dustin G. Mixon. Auto-tuning unit norm frames. *Applied and Computational Harmonic Analysis*, 32(1):1–15, 2012.
- [9] Peter G. Casazza and Jelena Kovačević. Equal-norm tight frames with erasures. *Advances in Computational Mathematics*, 18(2):387–430, February 2003.
- [10] Peter G. Casazza and Gitta Kutyniok, editors. *Finite Frames: Theory and Applications*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.
- [11] Ole Christensen. *An Introduction to Frames and Riesz Bases*. Birkhäuser, Cham, 2 edition, 2016.
- [12] Ken Dykema and Nate Strawn. Manifold structure of spaces of spherical tight frames. *International Journal of Pure and Applied Mathematics*, 28(2):217–256, 2006.
- [13] Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- [14] Martin Ehler. Random tight frames. *Journal of Fourier Analysis and Applications*, 18(1):1–20, 2012.
- [15] William Cole Franks and Ankur Moitra. Rigorous guarantees for tyler’s m-estimator via quantum expansion. In *Conference on Learning Theory*, pages 1601–1632. PMLR, 2020.
- [16] Vivek K. Goyal, Jelena Kovačević, and Jonathan A. Kelner. Quantized frame expansions with erasures. *Applied and Computational Harmonic Analysis*, 10(3):203–233, 2001.

- [17] Vivek K Goyal, Martin Vetterli, and Nguyen T Thao. Quantized overcomplete expansions in $\mathbb{R}^n$: analysis, synthesis, and algorithms. *IEEE Transactions on Information Theory*, 44(1):16–31, 1998.
- [18] Linus Hamilton and Ankur Moitra. The Paulsen problem made simple. *Israel Journal of Mathematics*, 246:299–313, 2021.
- [19] Patrick Hayden, Debbie W. Leung, and Andreas Winter. Aspects of generic entanglement. *Communications in Mathematical Physics*, 265(1):95–117, 2006.
- [20] Roderick B. Holmes and Vern I. Paulsen. Optimal frames for erasures. *Linear Algebra and its Applications*, 377:31–51, 2004.
- [21] Jelena Kovačević and Amina Chebira. Life beyond bases: The advent of frames (Part I). *IEEE Signal Processing Magazine*, 24(4):86–104, July 2007.
- [22] 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*, STOC 2018, pages 182–189, New York, NY, USA, 2018. Association for Computing Machinery.
- [23] Tsz Chiu Kwok, Lap Chi Lau, and Akshay Ramachandran. Spectral analysis of matrix scaling and operator scaling. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1184–1204. IEEE, 2019.
- [24] Béatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, 28(5):1302–1338, October 2000.
- [25] Michel Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [26] Yurii Lyubarskii and Roman Vershynin. Uncertainty principles and vector quantization. *IEEE Transactions on Information Theory*, 56(7):3491–3501, July 2010.
- [27] Simon Matabaigne, P.-A. Absil, and Nina Miolane. Bounds on the geodesic distances on the Stiefel manifold for a family of Riemannian metrics. *Linear Algebra and its Applications*, 730:1–34, 2026.
- [28] Dustin G Mixon, Tom Needham, Clayton Shonkwiler, and Soledad Villar. Three proofs of the Benedetto–Fickus theorem. In *Sampling, Approximation, and Signal Analysis: Harmonic Analysis in the Spirit of J. Rowland Higgins*, pages 371–391. Springer, 2024.
- [29] Tom Needham and Clayton Shonkwiler. Symplectic geometry and connectivity of spaces of frames. *Advances in Computational Mathematics*, 47(1):5, 2021.
- [30] Tom Needham and Clayton Shonkwiler. Toric symplectic geometry and full spark frames. *Applied and Computational Harmonic Analysis*, 61:254–287, 2022.
- [31] Du Nguyen. Curvatures of Stiefel manifolds with deformation metrics. *Journal of Lie Theory*, 32(2):563–600, 2022.
- [32] Akshay Ramachandran. *Geodesic Convex Analysis of Group Scaling for the Paulsen Problem and the Tensor Normal Model*. Phd thesis, University of Waterloo, November 2021.

- [33] Joel A. Tropp. A comparison principle for functions of a uniformly random subspace. *Probability Theory and Related Fields*, 153(3–4):759–769, 2012.
- [34] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.
- [35] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Yonina C. Eldar and Gitta Kutyniok, editors, *Compressed Sensing: Theory and Applications*, chapter 5, pages 210–268. Cambridge University Press, Cambridge, 2012.
- [36] Roman Vershynin. High-dimensional probability. *University of California, Irvine*, 10(11):31, 2020.

Samuel A. Ballas

Department of Mathematics, Florida State University
 sballas@fsu.edu

Ferhat Karabatman

Department of Mathematics, Florida State University
 fk22@fsu.edu

Tom Needham

Department of Mathematics, Florida State University
 tneedham@fsu.edu

A Numerical Experiments

We numerically illustrate Theorem 2.1 by sampling independent vectors uniformly from the sphere $\mathbb{S}^{d-1}_{\sqrt{d/n}}$ and estimating the probability $\mathbb{P}(\|S - I_d\|_{\text{op}} \geq \varepsilon)$ via Monte Carlo simulation, as shown in Figure 1. The dashed orange curve represents the theoretical upper bound in Theorem 2.1, while the solid blue curve represents the empirical probability. In all cases, the empirical probability decays rapidly as n increases, illustrating the concentration of the random frame operator around the identity. The theoretical bound captures the same exponential decay behavior, although it is conservative, as expected from a non-asymptotic concentration inequality.

We also illustrate the corresponding concentration result for the ball model. In Figure 2, the vectors are sampled uniformly from the ball of radius $\sqrt{(d+2)/n}$, which is chosen so that $\mathbb{E}S = I_d$. As in the sphere case, the empirical probability $\mathbb{P}(\|S - I_d\|_{\text{op}} \geq \varepsilon)$ decays rapidly as n increases. The theoretical bound captures the same decay behavior, while remaining conservative.

Finally, we numerically illustrate Proposition 3.10 by sampling matrices uniformly from the Stiefel manifold $\text{St}(n, d)$ and estimating the probability $\mathbb{P}\left(\max_{i \in \{1, \dots, n\}} \|u_i\| - \sqrt{\frac{d}{n}} \geq \varepsilon\right)$. Figure 3 compares the empirical probability with the theoretical upper bound.

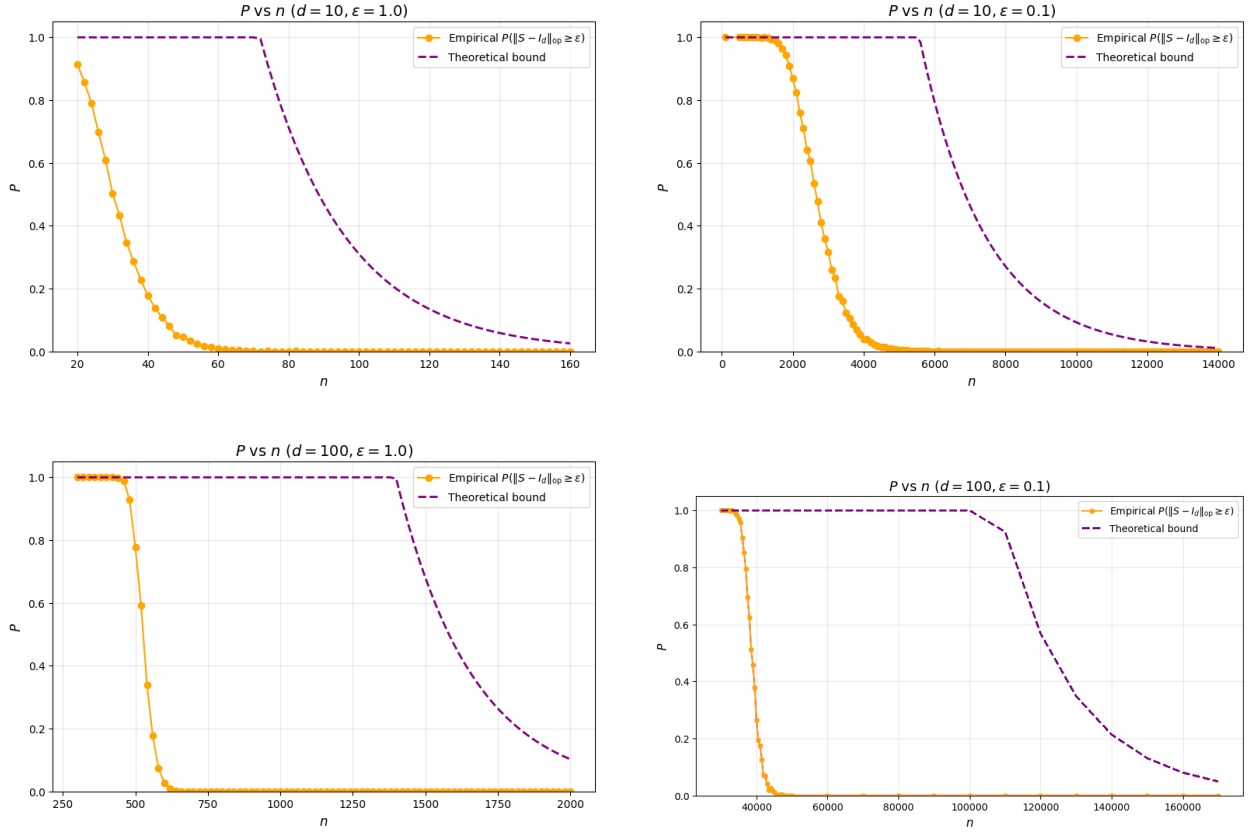


Figure 1: Empirical probabilities of $\|S - I_d\|_{\text{op}} \geq \varepsilon$ for the sphere model, compared with the bound in Theorem 2.1

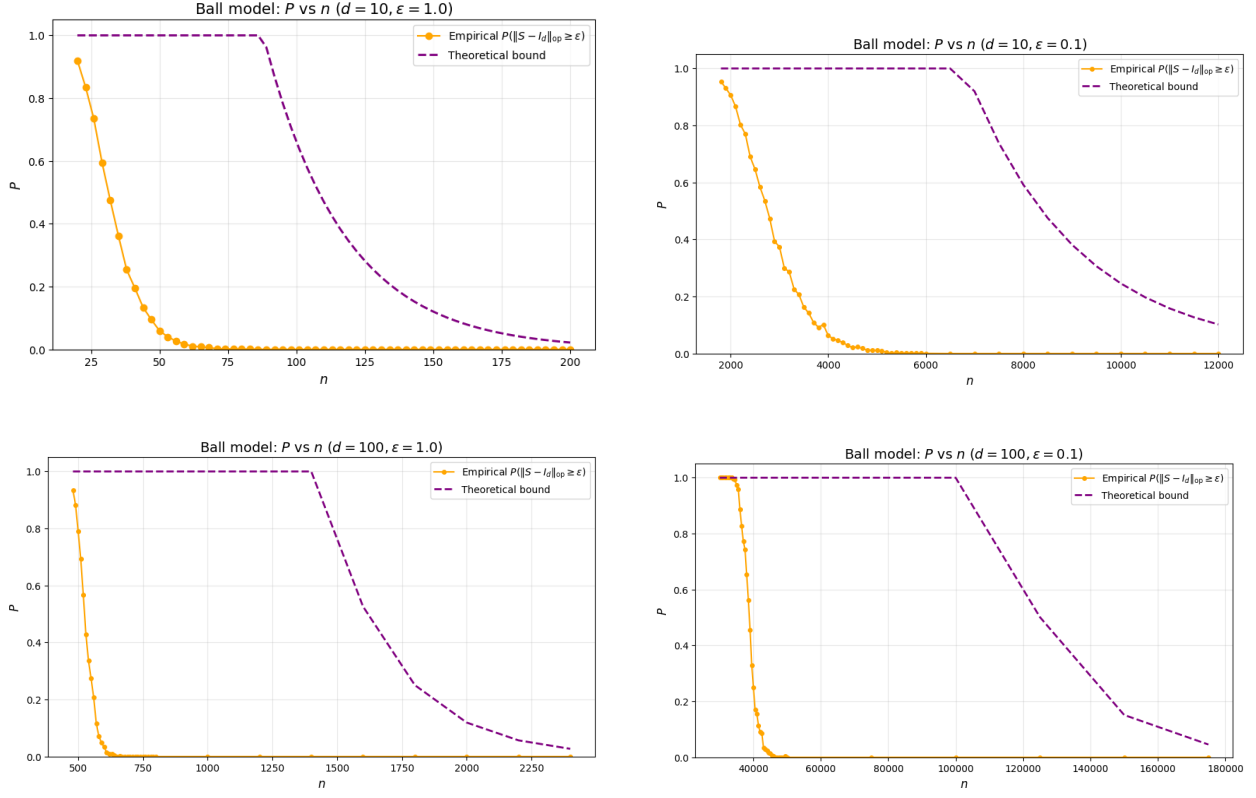


Figure 2: Empirical probabilities of $\|S - I_d\|_{\text{op}} \geq \varepsilon$ for the ball model, compared with the bound in Theorem 2.4

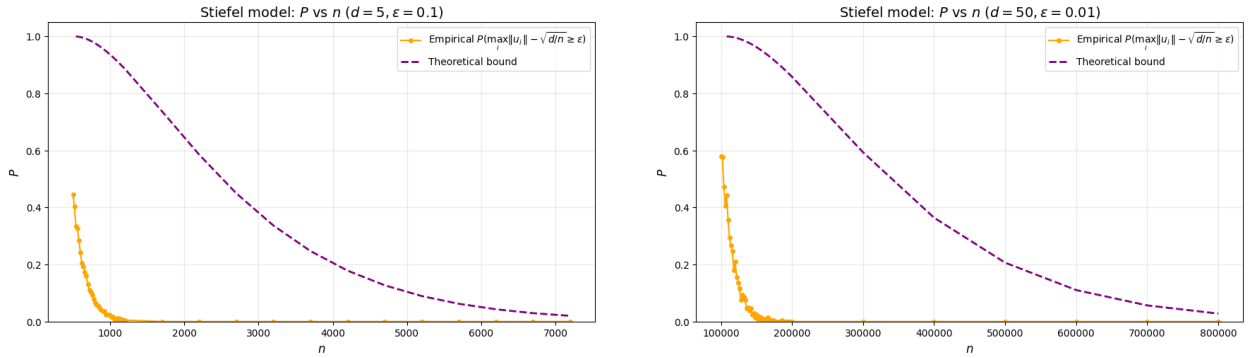


Figure 3: Empirical probabilities for the Stiefel model, compared with the bound in Proposition 3.10.

In conclusion, our numerical experiments support the theoretical concentration bounds presented in the paper. However, the empirical behavior suggests that there may be room for improvement in the concentration inequalities. We leave this as a potential future research direction.