Conic Frames, Cosine Measure, and Gaussian Frames

Jonathan Forstater JFORSTATER@UCDAVIS.EDU
Department of Mathematics
University of California, Davis
Davis, CA 95616-5270, USA

Thomas Strohmer STROHMER@MATH.UCDAVIS.EDU
Department of Mathematics
University of California Davis
Davis, CA 95616-5270, USA

Abstract

Conic combinations play an important role across mathematics in areas such as optimization, non-negative matrix factorization, and sparse coding. Frames which span their underlying space using only conic combinations, known as positive spanning sets or conic frames, are of special interest in derivative free optimization. In particular, sets with large cosine measure, a measure of how spread out the vectors are, are favored. We focus on the asymptotics of the cosine measure of Gaussian frames and present tail bounds in a variety of regimes.

1. Introduction

Frames are a fundamental concept in mathematics, particularly in areas such as harmonic analysis, linear algebra, functional analysis, signal processing, and numerical analysis [42, 12, 9]. They generalize the idea of a basis in a vector space, offering a more flexible and robust way to represent and analyze signals or data. Frames offer a powerful mathematical framework for understanding and manipulating data in high-dimensional spaces, providing a bridge between the rigid structure of bases and the flexibility needed for real-world applications. As such, they have found widespread applications in signal and image processing (e.g., wavelet frames for multiresolution analysis, Gabor frames for time-frequency analysis), data compression, machine learning, and numerical methods for solving partial differential equations.

Unlike a basis, a frame can be redundant, meaning its vectors are not necessarily linearly independent. Indeed, this non-uniqueness of representation coefficients turns out to be very useful for certain applications where it is desirable to restrict the coefficients to be *non-negative*, resulting in what is known as a *conic combination*. This type of constraint is fundamental to the mathematical field of optimization and has applications in many scientific and engineering problems, such as Non-Negative

Matrix Factorization, Sparse Coding and Dictionary Learning, and Derivative Free Optimization.

We organize the paper as follows. In Section 2, we review the necessary background on frames, convex geometry, and their applications. Then, in Section 3, we develop geometric characterizations of conic frames, establish their connection with bounded polyhedra, and present randomized and deterministic methods to make certain frames conic. In Section 4, we examine the relationship between coherence and the cosine measure and show that twin equiangular tight frames and regular polytopes need not maximize the cosine measure. Last, in Section 5, we identify the cosine measure with the minimum facet height of the convex hull of the frame and use this connection to establish semi-asymptotic formulas and concentration bounds for normalized Gaussian frames. We conclude by providing non-asymptotic tail bounds in the subexponential regime.

2. Mathematical Background

2.1 Frames

We recommend [9] as a great introduction to frame theory. A frame in a separable Hilbert space $\mathcal{H}$ is a sequence of vectors $\{f_k\}_{k\in I}$ (where I is an index set, possibly infinite) such that there exist constants $A > 0$ and $B < \infty$ satisfying the frame bounds for all vectors $f \in \mathcal{H}$,

$$A\|f\|^2 \leq \sum_{k\in I} |\langle f, f_k\rangle|^2 \leq B\|f\|^2. \tag{1}$$

Each frame has the associated frame operator, $S : \mathcal{H} \to \mathcal{H}$:

$$Sf = \sum_{k\in I} \langle f, f_k\rangle f_k.$$

Note that the frame operator is self-adjoint and the lower frame bound implies that S is (strictly) positive and thus invertible. This inverse allows us to define the canonical dual frame, $\{S^{-1}f_k\}_{k\in I}$, leading to the following two representations:

$$f = \sum_{k\in I} \langle f, f_k\rangle S^{-1}f_k = \sum_{k\in I} \langle f, S^{-1}f_k\rangle f_k$$

If the frame is tight, S and S^{-1} are both multiples of the identity, meaning that computing these representations are efficient.

In this paper we focus on *finite frames*, i.e., frames for a finite-dimensional Hilbert space [9, 39, 8]. In particular, we are interested in the case $\mathcal{H} = \mathbb{R}^n$. Thus, we consider vectors $f_1, \ldots, f_m$ with $m \geq n$. The upper frame bound is trivially satisfied in this case. To satisfy the lower frame bound in (1) the vectors $f_1, \ldots, f_m$ need to span

$\mathbb{R}^n$. Equivalently, we can require that the $n \times m$ matrix $[f_1, \ldots, f_m]$ has rank n. By a slight abuse of notation, we also write F for the matrix representing a frame, i.e., $F = [f_1, \ldots, f_m]$.

For certain applications, we are interested in how 'spread out' the frame vectors are. This is quantified by the notion of *coherence*:

Definition 2.1 (Coherence)**.** *The coherence, μ, of a frame F consisting of all non-zero vectors is:*

$$\mu(F) = \max_{k \neq l} \frac{|\langle f_k, f_l \rangle|}{\|f_k\| \|f_l\|}$$

Notice that coherence is equivalent to the minimum acute angle between the lines spanned by the frame vectors. By minimizing the coherence, the 'spread' of the vectors is thus maximized. The following lower bound on the coherence was originally shown by L. R. Welch [40]:

Theorem 2.2 (Welch Bound)**.** *Let $\{f_k\}_{k=1}^m$ be a unit norm frame for $\mathbb{R}^n$ or $\mathbb{C}^n$, then*

$$\mu \geq \sqrt{\frac{m-n}{n(m-1)}}$$

Tightness of a frame is a necessary condition to achieve the Welch bound, but it is not sufficient; this leads us to our next definition

Definition 2.3 (Equiangularity)**.** *A frame is called equiangular if $|\langle f_k, f_l \rangle| = \alpha$ for some fixed α and for all $k \neq l$.*

This leads us to *equiangular tight frames* (ETF), which rose to prominence via [34]. It is proven there that equality holds in the Welch bound if and only if a frame is an ETF.

2.2 Conic Frames and Convex Geometry

Conic frames are frames that can represent every vector in their underlying space as a conic combination (a linear combination with non-negative coefficients).

Definition 2.4 (Conic Frame)**.** *Let $F \in \mathbb{R}^{m \times n}$ be a finite frame such that for all $y \in \mathbb{R}^n$ there exists an $x \in \mathbb{R}^m$ with $x \geq 0$ such that $Fx = y$. Then F is a* conic frame.

These are fundamental to convex geometry, along with affine (coefficients sum to one) and convex combinations (both affine and conic).

A natural way to study frames is by studying the properties of their convex hulls, which are known as convex polytopes. In order to do so, we will first provide some background definitions and results from convex geometry.

Let $M \subseteq \mathbb{R}^n$. The affine hull (or affine span) of M, $\operatorname{aff} M$, is the smallest affine set that contains M. Similarly, the convex hull of M, $\operatorname{conv} M$ is the smallest convex set containing M.

Affine sets are in fact translates of linear subspaces (also known as a *flats*). That is, if A is an affine set then $A = x + L$, where $x \in \mathbb{R}^n$ and L is a linear subspace. The dimension of an affine set is defined as the dimension of L, while the dimension of a convex set, C, is the dimension of $\operatorname{aff} C$.

The following lemma (the proof of which is left to the reader) relates the affine span of a set to its linear span.

Lemma 2.5. *For any subset M of $\mathbb{R}^n$, if $0 \in \operatorname{aff} M$, then $\dim(\operatorname{aff} M) = \dim(\operatorname{span} M)$. Otherwise, $\dim(\operatorname{aff} M) = \dim(\operatorname{span} M) - 1$*

Convex sets that have a lower dimension than their ambient space have an empty interior. This leads us to the notion of *relative interior* which is guaranteed to exist for any non-empty convex set:

Definition 2.6. *The* relative interior *of a convex set C, $\operatorname{ri} C$ is the interior of C in $\operatorname{aff} C$.*

Note that a non-empty convex set in $C \subseteq \mathbb{R}^n$ has a non-empty interior if and only if $\operatorname{aff} C = \mathbb{R}^n$ and in such a case $\operatorname{ri} C = \operatorname{int} C$ [7].

As mentioned previously, a type of convex set that is particularly useful in understanding finite frames is the *convex polytope*, which is the convex hull of a finite set of points. Note that a convex polytope is closed and bounded and is thus a compact convex set. From here on, we may refer to convex polytopes simply as polytopes. The following lemma (which the reader can readily verify) allows us to characterize the relative interior of a polytope.

Lemma 2.7. *Let $P = \operatorname{conv}\{x_1, \ldots, x_m\}$ be a polytope in $\mathbb{R}^n$. $x \in \operatorname{ri} P$ if and only if x is a convex combination of $\{x_1, \ldots, x_m\}$ with strictly positive coefficients.*

In order to understand the structure of a polytope, we can look at its *faces*. A *supporting hyperplane* of a convex set is a hyperplane which contains the convex set in one of its associated halfspaces and intersects the set; a face is precisely this intersection.

An important feature of the faces of polytopes is that they themselves are polytopes. The following is a rephrasing of Proposition 2.3 in [43]:

Theorem 2.8 (Face of Convex Polytope is Convex Polytope)**.** *Let P be a convex polytope, then each face of P is itself a convex polytope. Specifically, if $A = \{x_1, ..., x_k\}$ and $P = conv(A)$, a face of P has the form $\operatorname{conv}(B)$ where $B \subseteq A$.*

Noting that a hyperplane is $(n-1)$-dimensional and every face is a subset of some hyperplane, we have the following corollary:

Corollary 2.9 (Faces of Polytopes are at most $n-1$-dimensional)**.** *Let K be an n-dimensional polytope embedded in $\mathbb{R}^n$, then each face of K is a k-dimensional polytope (with $k < n$), defined by a subset of the points that define K.*

The following two theorems give conditions for the existence of supporting hyperplanes. They can be found in Sections 2.5.1 and 2.5.2 of [5] respectively where they are phrased slightly differently.

Theorem 2.10 (Separating Hyperplane Theorem)**.** *Let C and D be non-empty disjoint convex sets in $\mathbb{R}^n$. Then there is a hyperplane H such that $C \subseteq H^+$ and $D \subseteq H^-$.*

Theorem 2.11 (Supporting Hyperplane Theorem)**.** *Let C be a non-empty closed convex set in $\mathbb{R}^n$ and let b be a boundary point of C. Then there exists at least one supporting hyperplane of C containing b.*

This last theorem can be restated in a way that will be useful later in our study of conic frames:

Corollary 2.12. *Each boundary point of a closed convex set C in $\mathbb{R}^n$ lies in at least one face of C.*

2.3 Previous Work on Conic Combinations and their Applications

Chandler Davis' paper [13] is one of the starting points for studying conic combinations[1]. In that paper, he refers to conic combinations as positive combinations and defines notions of positive independence, positive span, and positive basis, which mirror their linear versions. He then mainly focuses on discussing positive bases, which are guaranteed to have between $n+1$ and $2n$ elements in $\mathbb{R}^n$. Follow-up work by [26] generalizes some results to infinite dimensional spaces and develops the theory of positive bases further.

At this point, the term *positive spanning set* came into use to refer to the object we now refer to as a conic frame. In [31], Regis discusses positive spanning sets, positive independent sets, and positive bases further. It also shows how many of the things we take for granted about linear versions of such sets are not necessarily true for the positive versions. He also discusses algorithms for extending any set of vectors in a linear subspace to a positive spanning set of that subspace by adding additional vectors. In Section 3, we present deterministic and randomized algorithms to turn a full-spark frame into a conic frame by merely changing the signs on the vectors of the frame without adding vectors.

Conic combinations appear in a wide range of scientific and engineering disciplines due to their role in describing positive linear systems, optimization problems, and

1. The term *frame* also appears in the paper by Chandler Davis, however, it has a completely different meaning there and should not be confused with the widely established concept of a frame as we use it in this paper

geometric structures. In particular, conic combinations underpin the theory of convex cones, which are central to many areas of applied mathematics. A fundamental example is found in linear programming and its generalization, conic programming, where feasible regions are often defined using conic constraints—sets closed under conic combinations [5].

Non-negative Matrix Factorization (NMF) is a prime example where conic combinations are implicitly central [25]. NMF decomposes a non-negative data matrix into two non-negative matrices. The columns of the data matrix are approximated as conic combinations of the columns of one of the factor matrices (the "basis vectors" or "features").

In Audio Source Separation [38], mixed audio signals are decomposed into individual sound sources using non-negative combinations of atoms with NMF techiniques [37] or probability distributions in a technique known as probabilistic latent component analysis (PLCA) [33].

In Beamforming [36], an array of antennas is arranged so that by changing the phase and relative amplitude of the signal sent from each antenna, the signal can be formed into a directed beam. Thus, the signal can be viewed as coefficients in a linear system. In general, the coefficients can be complex-valued, but by reformulating this system in terms of intensity, the parameters become non-negative real numbers, leading to a conic combination [17].

Conic combinations also play an important role in finance in portfolio theory. When constructing investment portfolios, one often considers non-negative weights for different assets. Imposing such a condition can reduce risk in estimated optimal portfolios [22].

Another domain where conic combinations are relevant is machine learning, particularly in areas such as sparse coding / dictionary learning [1] and support vector machines (SVMs). In sparse coding, an over-complete dictionary of atoms is learned which can be used to represent signals by means of a sparse linear combination of these atoms. Hoyer [21] approximates nonnegative data using learned nonnegative dictionary atoms and nonnegative coefficients, incorporating a sparsity penalty to encourage representations with few active atoms. As for SVMs, Potluru et al. [30] reduce totally nonnegative least squares to a single-class SVM, relating the support of the nonnegative coefficient vector to the SVM support vectors.

In economics and game theory, conic combinations help model production sets or strategy spaces where only nonnegative actions or resources are meaningful. This modeling is especially prevalent in input-output analysis and cost modeling, where outputs depend linearly on inputs with nonnegative coefficients [32].

But perhaps one of the most natural applications of conic combinations and conic frames in particular is Derivative-free Optimization (DFO). In many real-life situations, gradient information is unavailable or costly to compute. In these cases a DFO method, such as directional direct search (DDS) is particularly useful. In DDS [24], a descent direction is found by polling objective values in a number of directions. These

directions correspond to a positive spanning set, or conic frame as it is called here. Under mild assumptions, directional direct search is guaranteed to converge [10]. In proving convergence guarantees, the cosine measure of a conic frame is of particular importance.

Definition 2.13 (Cosine Measure)**.** *The cosine measure of a set of nonzero vectors, F, is defined as*

$$\operatorname{cm} F = \min_{0 \neq v \in \mathbb{R}^n} \max_{f \in F} \frac{v^\top f}{\|v\| \|f\|}$$

The cosine measure of a PSS is guaranteed to fall between 0 and 1 and PSSs with maximal cosine measure are often desired for convergence purposes. In [28], Nævdal finds positive bases with $n+1$ and $2n$ elements with maximal cosine measure, but for the vast majority of conic frames in $\mathbb{R}^n$ the question of which set of m elements has maximal cosine measure remains open [14].

3. Existence and construction of conic frames

With the above definitions and results in hand, we are ready to prove some results about conic frames. We begin with necessary and sufficient conditions for a frame to be a conic frame and then connect conic frames with bounded polyhedra. Finally, we present methods to construct a conic frame *de novo* or from an existing full-spark frame with high probability or deterministically.

3.1 Conic Frames and the Origin

Some of the results in this section have been shown previously by Regis [31] and Davis [13], but we present new proofs using polytopes to provide geometric intuition. We begin by demonstrating that the interior of the convex hull of a frame containing the origin is equivalent to certain conditions.

Lemma 3.1. *Let $F = \{x_1, \ldots, x_m\}$ be a frame in $\mathbb{R}^n$ and let $P = \operatorname{conv}(F)$. $0 \in \operatorname{int}(P)$ if and only if 0 can be written as a strictly positive convex combination of $x_1, \ldots, x_m$.*

Proof. ($\implies$) Assume $0 \in \operatorname{int} P$. Then $\operatorname{int} P \neq \emptyset$, so $0 \in \operatorname{ri} P = \operatorname{int} P$ and, by Lemma 2.7, 0 can be written by a strictly positive convex combination of the frame vectors.

($\impliedby$) Let 0 be representable as a strictly positive convex combination of the frame vectors. Then by Lemma 2.7, $0 \in \operatorname{ri} P$. It suffices to show that $\operatorname{aff} P = \mathbb{R}^n$ or equivalently $\dim \operatorname{aff} P = n$. Note $0 \in \operatorname{ri} P \subseteq \operatorname{aff} P$. Thus by Lemma 2.5 we have $\dim(\operatorname{aff} P) = \dim(\operatorname{span} P) = n$, where the last equality follows since finite frames are spanning sets. □

Lemma 3.2. *Let F be a frame in $\mathbb{R}^n$ and $P = \operatorname{conv}(F)$. Then $0 \notin \operatorname{int} P$ if and only if there is a hyperplane passing through 0 such that F is contained in one of the resulting halfspaces.*

Proof. ($\implies$) Assume $0 \notin \operatorname{int} P$. Then, we have two cases:

1. If $0 \notin P$, Theorem 2.10 gives us a separating hyperplane that contains F in one of the resulting halfspaces. If necessary, take the parallel hyperplane that contains 0.

2. $0 \in \operatorname{bd} P$, so by Corollary 2.12, 0 lies in some face of P, say $P \cap H$. In which case, H is the desired hyperplane.

($\impliedby$) Without loss of generality, assume there is a hyperplane, H, passing through 0 such that $F \subseteq H^+$. Since halfspaces are convex, $\operatorname{conv} F \subseteq H^+$. Thus $\operatorname{conv} F$ cannot contain a neighborhood of the origin, that is, $0 \notin \operatorname{int} P$. □

We now prove that we can characterize conic frames by inspecting the interior of their convex hull.

Theorem 3.3. *Let F be a frame in $\mathbb{R}^n$ and let $P := \operatorname{conv} F$. $0 \in \operatorname{int} P$ if and only if we can write any vector in $\mathbb{R}^n$ as a conic combination of the vectors that define F.*

In the proof we will use Farkas' Lemma, a well-known result in optimization:

Lemma 3.4 (Farkas' Lemma [43])**.** *Let $A \in \mathbb{R}^{n \times m}$ and $z \in \mathbb{R}^n$. Then exactly one of the following is true:*

1. *There exists a point $x \in \mathbb{R}^m$ with $Ax = z$ and $x \geq 0$*

2. *There exists a vector $d \in \mathbb{R}^n$ with $A^\top d \geq 0$ and $d^\top z < 0$*

Proof of Theorem 3.3. We proceed by proving the negation of each statement implies the other.

($\implies$) Assume that there is a vector $z \in \mathbb{R}^n$ that cannot be written as a conic combination of the vectors of F, so assertion 1 of Lemma 3.4 is false. Then assertion 2 of Lemma 3.4 must be true, and thus there is a $d \in \mathbb{R}^n$ with $F^\top d \geq 0$. Let $H := \{x \in \mathbb{R}^n \mid \langle x, d \rangle = 0\}$. Then $0 \in H$ and $F \subseteq H^+$, so by Lemma 3.2, $0 \notin \operatorname{int} P$.

($\impliedby$) Now assume $0 \notin \operatorname{int} P$, so by Lemma 3.2 there is a hyperplane passing through the origin such that F is contained in one of its halfspaces. That is, there is a unit vector $u \in \mathbb{R}^n$ with $\langle f_i, u \rangle \geq 0$ for all i and thus $F^\top u \geq 0$. But then assertion 2 of Lemma 3.4 is true for $z = -u$ since $(-u)^\top u = -1 < 0$, so assertion 1 of Lemma 3.4 must be false, i.e. $-u$ cannot be written as a conic combination of the vectors that define F. □

This proof is fairly concise, but it does not provide much intuition for why it is true. We thus present a different proof for the forward direction which allows us to say something stronger:

Theorem 3.5. *Let $F = \{f_1, \ldots, f_m\}$ be a frame in $\mathbb{R}^n$, and let $P = \operatorname{conv} F$. Then $0 \in \operatorname{int} P$ if and only if every vector in $\mathbb{R}^n$ is a conic combination of at most n vectors of F.*

The proof of this result relies on the fact that we can scale $\operatorname{conv} F$ so that any point lies on its boundary, and thus one of the faces of $\operatorname{conv} F$. This provides a scaling of a convex combination which results in a conic combination. We formalize this scaling in the following lemma:

Lemma 3.6 (Scaling Lemma)**.** *Let K be a compact convex set in $\mathbb{R}^n$ such that $0 \in \operatorname{int} K$ and let $x \neq 0$ be a point in $\mathbb{R}^n$. Then there exists an $r > 0$ such that x lies on the boundary of rK.*

Proof. Let $T := \{t \geq 0 : tx \in K\}$. Since K is closed, T is closed. Boundedness of K gives an $R > 0$ such that $t \in T \implies t\|x\| \leq R$, so T is bounded. Moreover, $0 \in \operatorname{int} K$ implies that T contains a positive number. Hence $M := \max T$ exists and is positive.

By maximality, $Mx \in K$, while $(M + h)x \notin K$ for every $h > 0$. These outside points approach Mx, so $Mx \in \partial K$. Taking $r = 1/M$, positive scaling preserves boundaries and therefore $x = r(Mx) \in r\,\partial K = \partial(rK)$. □

We now prove Theorem 3.5:

Proof. ($\implies$) Suppose $0 \in \operatorname{int} P$, and let $x \in \mathbb{R}^n$. If $x = 0$, use $0 = 0f_1$. Now assume $x \neq 0$.

By Lemma 3.6, there exists $r > 0$ such that $x \in \partial(rP)$. By Corollaries 2.9 and 2.12, x belongs to a proper face G of rP of dimension $k \leq n - 1$, where $G = \operatorname{conv}\{rf_i : i \in I\}$ for some $I \subseteq \{1, \ldots, m\}$.

Applying Carathéodory's theorem in the k-dimensional affine hull of G, there exist indices $i_1, \ldots, i_q \in I$, with $q \leq k + 1 \leq n$, and coefficients $\lambda_1, \ldots, \lambda_q \geq 0$ such that

$$x = \sum_{j=1}^{q} \lambda_j (r f_{i_j}), \qquad \sum_{j=1}^{q} \lambda_j = 1.$$

Therefore,

$$x = \sum_{j=1}^{q} (r\lambda_j) f_{i_j}$$

is a conic combination of at most n vectors of F.

($\impliedby$) Suppose every vector in $\mathbb{R}^n$ is a conic combination of at most n vectors of F. Then F is a conic frame, so $0 \in \operatorname{int} P$ by Theorem 3.3. □

Taking Lemma 3.1 and Theorem 3.5 together, we have the following:

Corollary 3.7. *Let F be a frame in $\mathbb{R}^n$ consisting of m vectors and let $P = \operatorname{conv} F$. We can write any vector in $\mathbb{R}^n$ as the conic combination of at most n of the vectors that define F if and only if 0 can be written as a strictly positive convex combination of the vectors of F.*

Corollary 3.7 gives us a simple method to check if a frame is conic: we just need to check if 0 can be represented by a strictly positive convex combination of the frame vectors. This is easily verified by solving the following feasibility problem:

$$\begin{aligned} \text{find} \quad & x \\ \text{s.t.} \quad & Fx = 0 \\ & x > 0 \\ & \mathbb{1}^\top x = 1 \end{aligned}$$

We note here that Theorem 3.6 in [13] shows that a frame is conic if and only if 0 can be written as a strictly positive combination of the vectors of the frame. Our proof provides a new perspective and intuition by looking at the polytope defined by a frame.

We summarize the results of this section in the following theorem.

Theorem 3.8. *Let $F = \{x_1, \ldots, x_m\}$ be a frame in $\mathbb{R}^n$ and let $P = \operatorname{conv}(F)$. The following are equivalent:*

(a) $0 \in \operatorname{int} P$

(b) 0 *can be written as a strictly positive convex combination of* $x_1, \ldots, x_m$

(c) There is no hyperplane passing through 0 *such that* F *is contained in one of its resulting halfspaces*

(d) Any vector in $\mathbb{R}^n$ *can be written as a conic combination of the vectors that define* F

(e) Any vector in $\mathbb{R}^n$ *can be written as the conic combination of at most* n *of the vectors that define* F

3.2 Conic Frames and Polyhedra

Recall that a polyhedron, C, is the intersection of finitely many closed halfspaces,

$$C = \{x \in \mathbb{R}^n : \langle s_j, x \rangle \leq r_j,\ j = 1, \ldots, m\},$$

where the s_j are nonzero. A bounded polyhedron is in fact a polytope. And as the following theorem shows, there is a one-to-one correspondence between bounded polyhedra and conic frames.

Theorem 3.9. *A nonempty polyhedron C is bounded if and only if its defining normals $s_1, \ldots, s_m$ form a conic frame for $\mathbb{R}^n$.*

Before proving this theorem, we introduce the notion of a normal cone as well as a characterization of the normal cone of a polyhedron which can be found in ([20] Example 5.2.6(b)).

Definition 3.10 (Normal Cone)**.** *The direction $s \in \mathbb{R}^n$ is said to be* normal *to C at $x \in C$ when $\langle s, y - x\rangle \leq 0$ for all $y \in C$. The set of all such directions is called the* normal cone *to C at x, denoted by $N_C(x)$.*

Using the definition of a polyhedron above, define $J(x) := \{j = 1, \ldots, m : \langle s_j, x\rangle = r_j\}$, the index-set of active constraints at $x \in C$. Then $N_C(x) = \{\sum_{j \in J(x)} \lambda_j s_j : \lambda_j \geq 0\}$.

Proof. ($\Longrightarrow$) Suppose C is bounded. Because it is an intersection of closed halfspaces, it is closed; hence it is compact.

Take any $v \in \mathbb{R}^n$. The continuous linear functional $x \longmapsto \langle v, x\rangle$ attains its maximum on C, say at $x_v \in C$. Therefore, for every $y \in C$, $\langle v, y\rangle \leq \langle v, x_v\rangle$, or equivalently, $\langle v, y - x_v\rangle \leq 0$. By the definition of the normal cone, this says $v \in N_C(x_v) = \text{cone}\{s_j : j \in J(x_v)\}$, where $J(x_v) = \{j : \langle s_j, x_v\rangle = r_j\}$. Consequently, there are nonnegative coefficients such that

$$v = \sum_{j \in J(x_v)} \lambda_j s_j.$$

Since v was arbitrary, $\text{cone}\{s_1, \ldots, s_m\} = \mathbb{R}^n$. Thus the normals positively span $\mathbb{R}^n$. They also span it linearly, so they form a finite frame, and therefore a conic frame.

($\Longleftarrow$) Suppose the normals form a conic frame. For each coordinate vector e_i, choose nonnegative coefficients a_{ij}, b_{ij} such that

$$e_i = \sum_{j=1}^{m} a_{ij} s_j, \qquad -e_i = \sum_{j=1}^{m} b_{ij} s_j.$$

For any $x \in C$, the defining inequalities then give

$$x_i = \sum_{j=1}^{m} a_{ij} \langle s_j, x\rangle \leq \sum_{j=1}^{m} a_{ij} r_j,$$

and similarly,

$$-x_i \leq \sum_{j=1}^{m} b_{ij} r_j.$$

Hence

$$-\sum_{j=1}^{m} b_{ij} r_j \leq x_i \leq \sum_{j=1}^{m} a_{ij} r_j \qquad (i = 1, \ldots, n).$$

Every coordinate is bounded above and below by fixed finite constants, so C is bounded.

□

3.3 Random Conic Frames

We now turn to randomly generating a conic frame with high probability. To do so, we begin by stating Wendel's Theorem, which provides the probability that a set of randomly generated vectors on the unit sphere all lie on some hemisphere.

Theorem 3.11 (Wendel's Theorem [41]). *Let $x_1, x_2, \dots x_m$ be random unit vectors in $\mathbb{R}^n$ whose joint distribution is invariant under all reflections through the origin, i.e., independently changing the signs of any subset of the vectors, and is such that with probability one all subsets of size n are linearly independent, then the probability, $p_{n,m}$, that all the vectors lie on some hemisphere is:*

$$p_{n,m} = 2^{-m+1} \sum_{k=0}^{n-1} \binom{m-1}{k}$$

Note that $\binom{m-1}{k} = 0$ for $m - 1 < k$, so $p_{n,m} = 1$ for $m \leq n$.

A hemisphere is determined by a hyperplane through the origin, so $p_{n,m}$ is equivalently the probability that there exists a hyperplane through the origin defining a halfspace that contains $x_1, \dots, x_m$.

Theorem 3.12 (Probability that a random frame is conic). *Let F be a frame consisting of m vectors in in $\mathbb{R}^n$ satisfying the hypotheses of Theorem 3.11, then F is a conic frame with probability $1 - p_{n,m}$.*

Proof. By Theorem 3.8, F is conic if and only if there is no hyperplane through the origin defining a halfspace that contains F, which by Theorem 3.11 is precisely $1 - p_{n,m}$. □

We note that a set of Gaussian vectors satisfies the hypotheses of Theorem 3.12 almost surely.

A natural question that arises is whether we can turn a given frame easily into a conic frame? There are at least two trivial ways to turn any frame $\{f_k\}_{k=1}^m$ into a conic frame: 1. Just add all negative elements $\{-f_k\}_{k=1}^m$. While this preserves the condition number of the frame operator, it doubles the number of elements in the frame, which may not be desirable. 2. Add $f_{m+1} := -\frac{1}{m}\sum_{k=1}^m f_k$ to the frame. A simple calculation shows that the resulting frame $\{f_k\}_{k=1}^{m+1}$ is conic. While this approach adds only one element, the condition number of the frame operator of this so augmented frame is worse.

A more interesting way to turn a frame into a conic frame is by switching the signs of some frame elements. This procedure neither increases the number of elements, nor does it change the frame bounds. But can switching some signs of a frame always yield a conic frame? Moreover, without further assumptions about the frame, a deterministic search through all signs is computationally not feasible if m is large. Is there an efficient randomized algorithm to do so?

For specific cases, we will provide both a deterministic and a randomized algorithm to turn a frame into a conic frame via sign changes.

The following theorem provides a method to turn a frame into a conic frame with high probability. We need only ensure that we apply a distribution to the frame vectors that satisfies the hypotheses for Theorem 3.12.

Theorem 3.13. *Let F be a frame in $\mathbb{R}^n$ such that every subset of n of its columns is linearly independent and let $B_1, B_2, \ldots, B_m$ be i.i.d. random variables such that $\mathbb{P}(B_i = 1) = \frac{1}{2}$ and $\mathbb{P}(B_i = -1) = \frac{1}{2}$. If b_i is a realization of B_i and B is the $m \times m$ diagonal matrix with $B_{ii} = b_i$, and G is the matrix with columns equal to the unit normalized columns of F then GB is a conic frame with probability $1 - p_{n,m}$.*

Proof. First note that normalization preserves conicity and linear independence. Then note that the joint distribution of the columns of GB are invariant under all reflections through the origin by construction. Thus FB satisfies the conditions of Theorem 3.11, so the result follows by Theorem 3.12. □

The reader may wonder how large does m have to be compared to n for this method to have a reasonable probability of success? The probability of success is exactly $1/2$ when $m = 2n$. It is not difficult to see that by Hoeffding's inequality, $1 - p_{n,m} \geq 1 - \delta$ whenever

$$m \geq 2n + 2\sqrt{n \log(1/\delta)} + 2\log(1/\delta),$$

while for $m \leq 2n - 1$ the probability of success becomes exponentially small.

Regarding dictionary learning, given a full spark [15, 2]) n-dimensional dictionary, i.e. every size n subset of its atoms is linearly independent, Theorem 3.13 can be used to make it conic. This method leaves the frame bound and sparsity of the system unchanged.

3.4 A Deterministic Construction by Sign Changes

In addition to our randomized method, we also present a deterministic method to make a full spark frame conic.

Theorem 3.14. *Let $F = [f_1, \ldots, f_m] \in \mathbb{R}^{n \times m}$ be full spark, with $m \geq n + 1$. A diagonal matrix D with diagonal entries in $\{-1, 1\}$ can be constructed deterministically such that FD is a conic frame.*

Proof. Since F is full spark, $A = [f_1, \ldots, f_n]$ is invertible. Solve

$$Ac = f_{n+1}.$$

Every c_i is nonzero: otherwise f_{n+1} would lie in the span of $n - 1$ of the first n columns, contradicting full spark.

Define

$$g_i = \operatorname{sgn}(c_i) f_i \quad (1 \le i \le n), \qquad g_{n+1} = -f_{n+1}.$$

Then

$$\sum_{i=1}^{n} |c_i| g_i + g_{n+1} = 0.$$

Thus these $n+1$ vectors span $\mathbb{R}^n$ and admit a dependence with strictly positive coefficients. so by Theorem 3.8, they positively span $\mathbb{R}^n$. Assign arbitrary signs to the remaining columns; adding vectors preserves positive spanning. □

This construction also preserves the frame bounds and sparsity and has complexity $O(n^3)$.

4. Maximization of Cosine Measure

Amidst studying conic frames and the cosine measure, we noticed the similarity between coherence for a frame and cosine measure. Noting that ETFs achieve minimal coherence, it seems obvious to conjecture that ETFs achieve maximal cosine measure. However, one major difference between cosine measure and coherence is that the formula for cosine measure does not include an absolute value. A natural fix for this is to consider instead a Twin-ETF, defined below.

Definition 4.1. *Let F be an Equiangular Tight Frame. Then the set of vectors in F and $-F$, denoted $[F, -F]$, is a Twin-ETF.*

Note that for a Twin ETF,

$$\max_{g \in [F,-F]} \langle v, g \rangle = \max_{f \in F} |\langle v, f \rangle|$$

for every v.

The other difference between minimal coherence and cosine measure is that cosine measure checks each vector in the underlying space, while coherence only looks at the frame vectors themselves. This lead us to conjecture that perhaps regular polytopes (polytopes with maximal symmetry) might maximize cosine measure. In particular, it seems intuitively clear that a regular polytope that is also a twin-ETF, an object that is optimal for several geometrically inspired measures, should also be optimal with respect to the cosine measure. Alas, we were surprised to find that neither of these conjectures turn out to be true. In fact, a counterexample for 8 vectors in $\mathbb{R}^3$ demonstrates this.

Begin by noting that the following frame is a unit norm ETF in $\mathbb{R}^3$.

$$F = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

Thus $[F, -F]$ is a Twin ETF. Additionally, it is not difficult to see that $[F, -F]$ is in fact the unit norm cube and is thus also a regular polytope. However, we will demonstrate shortly that this conic frame does not have the maximal cosine measure for 8 vector frames in $\mathbb{R}^3$. In hindsight this is less surprising, since the coherence of a frame is a *packing* quantity, while the cosine measure is a *covering* quantity.

The square-antiprism is similar to a cube, but the top and bottom faces are rotated 45° from each other. That is, for some value of the parameter $0 < a < 1$ with $r = \sqrt{1 - a^2}$, a square antiprism has the following form:

$$S = \begin{bmatrix} r\cos(0) & r\cos(\frac{\pi}{2}) & r\cos(\pi) & r\cos(\frac{3\pi}{2}) & r\cos(\frac{\pi}{4}) & r\cos(\frac{3\pi}{4}) & r\cos(\frac{5\pi}{4}) & r\cos(\frac{7\pi}{4}) \\ r\sin(0) & r\sin(\frac{\pi}{2}) & r\sin(\pi) & r\sin(\frac{3\pi}{2}) & r\sin(\frac{\pi}{4}) & r\sin(\frac{3\pi}{4}) & r\sin(\frac{5\pi}{4}) & r\sin(\frac{7\pi}{4}) \\ a & a & a & a & -a & -a & -a & -a \end{bmatrix}$$

For $a = 0.7$, we have the following:

$$S = \begin{bmatrix} 0.7141 & 0 & -0.7141 & 0 & 0.505 & -0.505 & -0.505 & 0.505 \\ 0 & 0.7141 & 0 & -0.7141 & 0.505 & 0.505 & -0.505 & -0.505 \\ 0.7 & 0.7 & 0.7 & 0.7 & -0.7 & -0.7 & -0.7 & -0.7 \end{bmatrix}$$

We implemented and ran the deterministic method from [19] in MATLAB to calculate the cosine measure of each of these conic frames. We obtained $\mathrm{cm}([F, -F]) \approx 0.5774$ while $\mathrm{cm}(S) \approx 0.6029$, meaning that this square antiprism has higher cosine measure than the unit norm cube. At first glance, this is surprising. However, in [14] it is shown that finding the set of m vectors in $\mathbb{R}^n$ with maximal cosine measure is equivalent to finding the most economical covering by S^{n-1} by m spherical caps. This problem is similar to (but still different from) the Tammes problem which looks for the spherical code of m points with largest angular separation between points. It is known that the solution for 8 points in $\mathbb{R}^3$ is the square antiprism [11].

But this is not the whole story. The best known 8-cap covering of the sphere is a D_2d point configuration [18] of the following form[2]:

$$F = \big\{(\pm s_1, 0, c_1),\ (0, \pm s_1, -c_1),\ (0, \pm s_2, c_2),\ (\pm s_2, 0, -c_2)\big\},$$
$$c_1 \approx 0.7867473615, \quad c_2 \approx 0.1715313650, \quad s_i = \sqrt{1 - c_i^2},$$

which has $\mathrm{cm}(F) \approx 0.6673188866$, an even higher cosine measure. Also note that the Tammes-optimal square anti-prism [16] has cosine measure $a \approx 0.511$, demonstrating that these problems are distinct.

2. As an abstract group, D_2d is isomorphic to the dihedral group of order 8.

5. The Cosine Measure of Gaussian Frames

5.1 Facet Height and Cosine Measure

We now provide an alternative characterization of the cosine measure of a polytope, namely the minimum facet height of the polytope.

Lemma 5.1 (Cosine measure as the minimum facet height)**.** *Let*

$$X = \{x_1, \dots, x_m\} \subset \mathbb{S}^{d-1}$$

and let

$$P := \operatorname{conv}(X).$$

Suppose that

$$0 \in \operatorname{int} P.$$

For each facet $F \in \mathcal{F}(P)$, let $u_F \in \mathbb{S}^{d-1}$ be its outward unit normal and let $H_F > 0$ be its height, so that

$$F \subseteq \{x \in \mathbb{R}^d : \langle u_F, x\rangle = H_F\}$$

and

$$P \subseteq \{x \in \mathbb{R}^d : \langle u_F, x\rangle \le H_F\}.$$

Then

$$\operatorname{cm} X = \min_{\|u\|=1} \max_{1\le i\le m} \langle u, x_i\rangle = \min_{F\in\mathcal{F}(P)} H_F.$$

Equivalently, the cosine measure is the radius of the largest Euclidean ball centered at the origin and contained in P.

Proof. Let

$$h_P(u) := \max_{x\in P} \langle u, x\rangle, \qquad u \in \mathbb{S}^{d-1},$$

be the support function of P. Because P is the convex hull of $x_1, \dots, x_m$, a linear functional attains its maximum over P at one of these points. Therefore,

$$h_P(u) = \max_{1\le i\le m} \langle u, x_i\rangle.$$

It follows that

$$\operatorname{cm} X = \min_{\|u\|=1} h_P(u). \tag{2}$$

Let

$$r_0(P) := \sup\left\{r \ge 0 : r\mathbb{B}_2^d \subseteq P\right\}$$

denote the radius of the largest Euclidean ball centered at the origin and contained in P. We first show that

$$r_0(P) = \min_{\|u\|=1} h_P(u). \tag{3}$$

For every $r \geq 0$, we claim that

$$r\mathbb{B}_2^d \subseteq P \quad \Longleftrightarrow \quad r \leq h_P(u) \quad \text{for every } u \in \mathbb{S}^{d-1}. \tag{4}$$

Suppose first that $r\mathbb{B}_2^d \subseteq P$. The support function of $r\mathbb{B}_2^d$ is

$$h_{r\mathbb{B}_2^d}(u) = \max_{\|x\| \leq r} \langle u, x \rangle = r, \qquad \|u\| = 1.$$

By monotonicity of support functions under set inclusion,

$$r = h_{r\mathbb{B}_2^d}(u) \leq h_P(u)$$

for every $u \in \mathbb{S}^{d-1}$.

Conversely, suppose that

$$r \leq h_P(u) \qquad \text{for every } u \in \mathbb{S}^{d-1},$$

but that $r\mathbb{B}_2^d \not\subseteq P$. Then there exists some $x \in r\mathbb{B}_2^d$ such that $x \notin P$. Since P is closed and convex, the strict separation theorem gives a unit vector $u \in \mathbb{S}^{d-1}$ satisfying

$$\langle u, x \rangle > \sup_{y \in P} \langle u, y \rangle = h_P(u).$$

On the other hand, because $\|x\| \leq r$, the Cauchy–Schwarz inequality gives

$$\langle u, x \rangle \leq \|u\| \, \|x\| \leq r \leq h_P(u),$$

which is a contradiction. This proves (4). Taking the largest admissible value of r in (4) gives (3).

It remains to express the centered inradius in terms of the facet heights. The facet half-space representation of P is

$$P = \bigcap_{F \in \mathcal{F}(P)} \left\{ x \in \mathbb{R}^d : \langle u_F, x \rangle \leq H_F \right\}. \tag{5}$$

For every facet F,

$$\max_{\|x\| \leq r} \langle u_F, x \rangle = r,$$

because $\|u_F\| = 1$. Consequently, (5) implies

$$\begin{aligned} r\mathbb{B}_2^d \subseteq P \quad &\Longleftrightarrow \quad r\mathbb{B}_2^d \subseteq \{x : \langle u_F, x \rangle \leq H_F\} \quad \text{for every } F \\ &\Longleftrightarrow \quad r \leq H_F \quad \text{for every } F \in \mathcal{F}(P). \end{aligned}$$

Therefore,

$$r_0(P) = \min_{F \in \mathcal{F}(P)} H_F. \tag{6}$$

Combining (2), (3), and (6), we conclude that

$$\operatorname{cm} X = \min_{\|u\|=1} h_P(u) = r_0(P) = \min_{F \in \mathcal{F}(P)} H_F.$$

□

5.2 Concentration of the Cosine Measure of Normalized Gaussian Frames

Besides providing interesting geometric intuition, the facet height characterization connects cosine measure to previous work by Bonnet and O'Reilly [4] on asymptotic properties of convex hulls of normalized Gaussian vectors. We adapt the facet-count estimates of [4, Theorem 3.4 and its proof in Section 6.2.1] to obtain quantitative bounds for the cosine measure. The lower-tail argument follows their Lemma 6.10, with a sharper binomial-coefficient bound.

Our contributions are the following:

1. We demonstrate semi-asymptotic exponential decay of the tail probabilities for sufficiently large dimension.

2. We provide a new chain of inequalities to correct an inaccuracy in their proof. Note: The conclusion of [4, Theorem 3.4] remains valid in the regimes of Theorem 5.2; only the stated bound in [4, Lemma 6.10] fails.

Theorem 5.2 (Exponential concentration of the cosine measure)**.** *Let $g_1, \dots, g_m$ be independent standard Gaussian vectors in $\mathbb{R}^n$, and define*

$$u_i := \frac{g_i}{\|g_i\|}, \qquad P_{n,m} := \operatorname{conv}\{u_1, \dots, u_m\}.$$

The vectors $u_1, \dots, u_m$ are independent and uniformly distributed on $\mathbb{S}^{n-1}$. Suppose that

$$n \to \infty \qquad \text{and} \qquad \frac{m}{n} \to \infty.$$

Then the following assertions hold for their cosine measure

$$\mu_{n,m} = \min_{\|v\|=1} \max_{1 \le i \le m} \langle u_i, v \rangle.$$

(Note we use $\mu_{n,m}$ to denote the cosine measure in order to emphasize the dependence on m and n.)

1. ***Subexponential regime.*** *If $\log m = o(n)$, then, for every fixed $\varepsilon > 0$, there exist constants $c, C > 0$ such that*

$$\mathbb{P}\left(\left| \frac{\mu_{n,m}}{\sqrt{2\log(m/n)/n}} - 1 \right| > \varepsilon \right) \le Ce^{-cn}$$

   for all sufficiently large n.

2. ***Exponential regime.*** *If*

$$\frac{\log m}{n} \longrightarrow \gamma \in (0, \infty),$$

then, for every fixed $\varepsilon > 0$, there exist constants $c, C > 0$ such that

$$\mathbb{P}\left(\left|\mu_{n,m} - \sqrt{1 - e^{-2\gamma}}\right| > \varepsilon\right) \leq Ce^{-cn}$$

for all sufficiently large n.

3. ***Superexponential regime.*** *If*

$$\frac{\log m}{n} \longrightarrow \infty,$$

then, for every fixed $\varepsilon \in (0, 1)$, there exist constants $c, C > 0$ such that

$$\mathbb{P}\left(\mu_{n,m} < 1 - \varepsilon\right) \leq C \exp\bigl(-c\, n \log m\bigr)$$

for all sufficiently large m, n.

Proof. For $n \geq 2$ and $m \geq n+1$, $P_{n,m}$ is full-dimensional almost surely. Following the signed-height convention of [4, Section 1], for every facet F, define its signed height $H_F \in [-1, 1]$ by its outward unit normal u_F:

$$F \subseteq \{x : \langle u_F, x\rangle = H_F\}, \qquad P_{n,m} \subseteq \{x : \langle u_F, x\rangle \leq H_F\}.$$

This definition applies whether or not the event

$$E_{n,m} := \{0 \in \operatorname{int} P_{n,m}\}$$

occurs.

By Lemma 5.1, on the event $E_{n,m}$,

$$\mu_{n,m} = \min_{F \in \mathcal{F}(P_{n,m})} H_F, \tag{7}$$

where $H_F > 0$ on $E_{n,m}$.

Define

$$h_1 = \sqrt{1 - \left(\frac{r_1 n (\log(m/n))^{3/2}}{m}\right)^{\frac{2}{n-1}}}$$

and

$$h_2 = \sqrt{1 - \left(\frac{r_2 n}{m}\right)^{\frac{2(n+1)}{(n-1)^2}}}, \tag{8}$$

where $r_1 > 0$ is sufficiently large and $r_2 > 0$ is sufficiently small; see [4, Theorem 3.4 and equation (6.24)]. Their notation uses n in place of our m, and d in place of our n. For all sufficiently large n, $0 < h_1 < h_2 < 1$.

Let

$$N_- := \#\{F : H_F \in [-1, h_1]\}$$

and

$$N_+ := \#\{F : H_F \in [h_2, 1]\}.$$

By the definition of $F_{[a,b]}$ in [4],

$$\mathbb{E}N_- = F_{[-1,h_1]}, \qquad \mathbb{E}N_+ = F_{[h_2,1]}. \tag{9}$$

We follow the facet-count argument in the proof of [4, Theorem 3.4], particularly the proofs of Lemmas 6.10 and 6.11. The strategy is to bound the expected number of facets whose heights fall outside the interval $[h_1, h_2]$. Markov's inequality then controls the probability that any such facet exists. On the event that the origin lies in the interior of the polytope, we use Lemma 5.1 to transfer this simultaneous control of facet heights to control of the cosine measure.

Repeating the proof of [4, Lemma 6.10] with a sharper binomial-coefficient estimate gives the following bound. For $\beta > -1$, let

$$c_\beta := \left(\int_{-1}^{1} (1 - t^2)^\beta \, dt \right)^{-1}.$$

Thus

$$c_\beta \int_{-1}^{1} (1 - t^2)^\beta \, dt = 1.$$

And set

$$L := \log(m/n), \qquad A := c_{\frac{n-3}{2}} \int_{h_1}^{1} (1 - t^2)^{\frac{n-3}{2}} \, dt.$$

The estimate from [4, Lemma 6.9] with $D = (n-3)/2$, together with the definition of h_1, gives,

$$A \geq \frac{c_{(n-3)/2}}{n-1} \frac{(1 - h_1^2)^{(n-1)/2}}{h_1} \left(1 - \frac{1 - h_1^2}{(n+1)h_1^2} \right).$$

By [4, equation (6.2)],

$$\frac{c_{(n-3)/2}}{n-1} \sim \frac{1}{\sqrt{2\pi n}}.$$

Moreover, [4, Lemma 6.8] gives $nh_1^2 \to \infty$, so the factor in parentheses tends to 1. Absorbing fixed numerical factors into $b_0 > 0$ gives:

$$A \geq b_0 \frac{(1 - h_1^2)^{(n-1)/2}}{h_1 \sqrt{n}}$$

Following the estimate in [4, equation (6.25)], write

$$h_1^2 = 1 - e^{-x}, \qquad x = \frac{2}{n-1} \log \left(\frac{m}{r_1 n L^{3/2}} \right).$$

Eventually $x > 0$ and $r_1 L^{3/2} \geq 1$. Using $1 - e^{-x} \leq x$,

$$h_1^2 \leq \frac{2}{n-1} \log\left(\frac{m}{r_1 n L^{3/2}}\right) \leq \frac{2L}{n-1}.$$

Therefore, for $n \geq 2$,

$$h_1 \sqrt{n} \leq \sqrt{\frac{2nL}{n-1}} \leq 2\sqrt{L}.$$

Substituting the definition of h_1 and this denominator bound gives

$$b_0 \frac{(1-h_1^2)^{(n-1)/2}}{h_1\sqrt{n}} \geq b_0 \frac{r_1 n L^{3/2}/m}{2\sqrt{L}} = \frac{b_0}{2} \frac{r_1 n L}{m}.$$

Setting $b_1 = b_0/2$, we have,

$$A \geq b_0 \frac{(1-h_1^2)^{(n-1)/2}}{h_1\sqrt{n}} \geq b_1 \frac{r_1 n L}{m}.$$

Using our notation, [4, equation (6.4)] states

$$F_{[a,b]} = 2\binom{m}{n} c_{\frac{n^2-2n-1}{2}} \int_a^b (1-h^2)^{\frac{n^2-2n-1}{2}} \left(c_{\frac{n-3}{2}} \int_{-1}^{h} (1-s^2)^{\frac{n-3}{2}}\, ds \right)^{m-n} dh.$$

The following chain of inequalities corrects the inaccuracy explained in Appendix B. It is inspired by [4, Lemma 6.11]. Bounding the inner normalized integral by $1 - A$ and the remaining normalized integral by 1 gives the first inequality below. Using $\binom{m}{n} \leq (em/n)^n$, $1 - A \leq e^{-A}$, and $m - n \geq m/2$ eventually,

$$\begin{aligned} \mathbb{E}N_- &\leq 2\binom{m}{n}(1-A)^{m-n} \\ &\leq 2\left(\frac{em}{n}\right)^n \exp\bigl(-br_1 nL\bigr) \\ &= 2e^n \left(\frac{m}{n}\right)^{-(br_1-1)n}, \end{aligned}$$

where $b = b_1/2 > 0$ is an absolute constant. Choose r_1 so that $c := br_1 - 1 > 0$. Because $L \to \infty$, eventually

$$\mathbb{E}N_- \leq e^{\log 2 + n - cnL} \leq e^{-(c/2)nL} = \left(\frac{m}{n}\right)^{-an}, \qquad a := c/2 > 0. \tag{10}$$

By [4, Lemma 6.11, equation (6.27)],

$$\mathbb{E}N_+ \leq \frac{C\sqrt{n}}{me}(r_2 e)^{n+1}. \tag{11}$$

We may choose r_2 sufficiently small that $r_2 e < 1$. Setting

$$c_2 := -\log(r_2 e) > 0,$$

equation (11) gives

$$\mathbb{E}N_+ \le C\frac{\sqrt{n}}{m}e^{-c_2 n}, \tag{12}$$

after changing the constant C.

Probability bounds. On the event $E_{n,m}$, equation (7) implies that $\mu_{n,m} < h_1$ forces $N_- > 0$, while $\mu_{n,m} > h_2$ forces $N_+ > 0$. On $E_{n,m}^c$, the facet half-space representation gives a facet with $H_F \le 0 < h_1$, so $N_- > 0$ as well. Hence, almost surely,

$$\{\mu_{n,m} \notin [h_1, h_2]\} \subseteq \{N_- > 0\} \cup \{N_+ > 0\}. \tag{13}$$

Since $E_{n,m}^c \subseteq \{N_- > 0\}$, the lower facet-count estimate also controls the probability that the frame fails to positively span $\mathbb{R}^n$. The union bound and Markov's inequality therefore give

$$\begin{aligned}\mathbb{P}\big(\mu_{n,m} \notin [h_1, h_2]\big) &\le \mathbb{P}(N_- > 0) + \mathbb{P}(N_+ > 0) \\ &\le \mathbb{E}N_- + \mathbb{E}N_+.\end{aligned} \tag{14}$$

Combining (10), (12), and (14), we obtain the quantitative cosine-measure estimate

$$\mathbb{P}\big(\mu_{n,m} \notin [h_1, h_2]\big) \le e^{-an\log(m/n)} + C\frac{\sqrt{n}}{m}e^{-c_2 n}. \tag{15}$$

Since $m/n \to \infty$, eventually $\log(m/n) \ge 1$, and hence

$$e^{-an\log(m/n)} \le e^{-an}.$$

The right-hand side of (15) is therefore at most

$$C' e^{-c' n} \tag{16}$$

for suitable constants $c', C' > 0$.

It remains to identify the deterministic asymptotics of the cutoffs h_1 and h_2. These follow from [4, Lemma 6.8], applied to the cutoffs as explained in their proof of Theorem 2.4.

Subexponential regime. By [4, Lemma 6.8(i) and the proof of Theorem 2.4],

$$\frac{h_i}{\sqrt{2\log(m/n)/n}} \longrightarrow 1, \qquad i = 1, 2. \tag{17}$$

Thus, for every fixed $\varepsilon > 0$, eventually

$$[h_1, h_2] \subseteq \left[(1-\varepsilon)\sqrt{\frac{2\log(m/n)}{n}}, (1+\varepsilon)\sqrt{\frac{2\log(m/n)}{n}}\right].$$

Combining this deterministic inclusion with (16) proves part (1).

Exponential regime. By [4, Lemma 6.8(ii)],

$$h_1, h_2 \longrightarrow \sqrt{1 - e^{-2\gamma}}. \tag{18}$$

Hence, for every fixed $\varepsilon > 0$, eventually

$$[h_1, h_2] \subseteq \left[\sqrt{1 - e^{-2\gamma}} - \varepsilon, \sqrt{1 - e^{-2\gamma}} + \varepsilon\right].$$

Combining this with (16) proves part (2).

Superexponential regime. Lemma 6.8(iii) of [4] gives

$$h_1 \longrightarrow 1. \tag{19}$$

Therefore, for every fixed $\varepsilon \in (0, 1)$, eventually $h_1 > 1 - \varepsilon$. The same argument as in (13), followed by (10), gives

$$\mathbb{P}(\mu_{n,m} < 1 - \varepsilon) \leq \mathbb{E}N_- \leq e^{-an \log(m/n)}. \tag{20}$$

Since $(\log m)/n \to \infty$,

$$\frac{\log n}{\log m} \to 0, \qquad \log(m/n) = \log m - \log n \sim \log m.$$

Consequently, eventually

$$e^{-an \log(m/n)} \leq e^{-(a/2)n \log m}.$$

This proves part (3). □

5.3 Non-asymptotic Gaussian Tail Bounds in the Subexponential Regime

In the preceding section, we proved semi-asymptotic convergence rates for the cosine measure of Gaussian frames in three regimes. We now provide non-asymptotic tail bounds for the cosine measure of Gaussian frames in the sub-exponential regime, beginning with a lower bound and then an upper bound. The proof techniques are very different from the ones used in the previous section.

5.3.1 A lower bound for the Gaussian cosine measure

Let

$$g_1, \ldots, g_m \overset{\text{iid}}{\sim} N(0, I_n), \qquad u_i := \frac{g_i}{\|g_i\|_2},$$

define

$$K_{n,m} := \min_{\|v\|_2 = 1} \max_{1 \leq i \leq m} \langle g_i, v \rangle,$$

and note the cosine measure,

$$\mu_{n,m} = \min_{\|v\|_2=1} \max_{1\le i\le m} \langle u_i, v\rangle .$$

Set

$$\alpha := \frac{m}{n}.$$

Theorem 5.3. *Fix $\varepsilon \in (0,1)$ and set*

$$\rho := \frac{\varepsilon}{3}, \qquad t := (1-\rho)\sqrt{2\log\alpha}, \qquad \beta := 1-(1-\rho)^2 = 2\rho-\rho^2.$$

There exist absolute constants $C, c > 0$ and a number $\alpha_0(\varepsilon)$ such that, whenever $\alpha \ge \alpha_0(\varepsilon)$,

$$\mathbb{P}\{K_{n,m} < t\} \le C\,\frac{\sqrt{\log\alpha}}{n\alpha^{\beta}} + 2e^{-cn}. \tag{21}$$

Moreover,

$$\begin{aligned}\mathbb{P}\left\{\mu_{n,m} < (1-\varepsilon)\sqrt{\frac{2\log(m/n)}{n}}\right\} \\ \le C\,\frac{\sqrt{\log(m/n)}}{n(m/n)^{\beta}} + 2e^{-cn} + m\exp\left(-\frac{n\varepsilon^2}{18}\right),\end{aligned} \tag{22}$$

where

$$\beta = \frac{2\varepsilon}{3} - \frac{\varepsilon^2}{9}.$$

In particular, if

$$\frac{m}{n} \to \infty, \qquad \log\frac{m}{n} = o(n),$$

then for every fixed $\varepsilon > 0$,

$$\mathbb{P}\left\{\mu_{n,m} < (1-\varepsilon)\sqrt{\frac{2\log(m/n)}{n}}\right\} \longrightarrow 0.$$

Proof. **Step 1: Gaussian min–max formulation.**

Let $G \in \mathbb{R}^{m\times n}$ be the matrix whose ith row is $g_i^{\top}$, and let

$$\Delta_m = \left\{ y \in \mathbb{R}^m : y_i \ge 0, \quad \sum_{i=1}^{m} y_i = 1 \right\}.$$

Since

$$\max_{1\le i\le m} x_i = \max_{y\in\Delta_m} y^{\top}x,$$

we have

$$K_{n,m} = \min_{v \in S^{n-1}} \max_{y \in \Delta_m} y^\top G v. \tag{23}$$

We use Gordon's one-sided Gaussian min–max theorem in the compact-set formulation of Thrampoulidis, Oymak, and Hassibi [35]. For our choice

$$\mathcal{S}_x = S^{n-1}, \qquad \mathcal{S}_y = \Delta_m, \qquad \psi \equiv 0,$$

it gives

$$\mathbb{P}\{K_{n,m} < s\} \le 2\,\mathbb{P}\{\mathcal{A}(g,h) \le s\}, \qquad s \in \mathbb{R}, \tag{24}$$

where $g \sim N(0, I_m)$ and $h \sim N(0, I_n)$ are independent, and

$$\mathcal{A}(g,h) = \min_{v \in S^{n-1}} \max_{y \in \Delta_m} \left\{ g^\top y + \|y\|_2 \, h^\top v \right\}.$$

Since the inner maximum is nondecreasing in the scalar $h^\top v$, and

$$\min_{v \in S^{n-1}} h^\top v = -\|h\|,$$

the outer minimum is attained at $v = -h/\|h\|$. Therefore

$$\mathcal{A}(g,h) = \max_{y \in \Delta_m} \{ g^\top y - \|h\| \, \|y\| \}.$$

Step 2: scalar characterization of the auxiliary problem.

Set

$$r := \|h\|_2 .$$

We consider

$$\mathcal{A}(g,h) = \max_{y \in \Delta_m} \left\{ g^\top y - r \, \|y\|_2 \right\}. \tag{25}$$

Let $y^\star$ be an optimizer. Since $y^\star \in \Delta_m$, we have $y^\star \neq 0$, so the Euclidean norm is differentiable at $y^\star$.

Introduce a Lagrange multiplier $\nu \in \mathbb{R}$ for the equality constraint

$$\sum_{i=1}^m y_i = 1$$

and multipliers $\gamma_i \ge 0$ for the inequality constraints

$$y_i \ge 0.$$

Using the Lagrangian

$$L(y, \nu, \gamma) = g^\top y - r \, \|y\|_2 - \nu \left(\sum_{i=1}^m y_i - 1 \right) + \sum_{i=1}^m \gamma_i y_i,$$

the stationarity conditions at $y^\star$ are

$$g_i - \nu + \gamma_i = r\,\frac{y_i^\star}{\|y^\star\|_2}, \qquad i = 1, \dots, m. \tag{26}$$

Complementary slackness gives

$$\gamma_i y_i^\star = 0. \tag{27}$$

If $y_i^\star > 0$, then $\gamma_i = 0$, and therefore

$$g_i - \nu = r\,\frac{y_i^\star}{\|y^\star\|_2}. \tag{28}$$

In particular, $g_i > \nu$ on every active coordinate.

If $y_i^\star = 0$, then (26) gives $g_i - \nu = -\gamma_i \le 0$, and hence $g_i \le \nu$. Thus,

$$y_i^\star > 0 \quad \Longleftrightarrow \quad g_i > \nu.$$

Consequently,

$$(g_i - \nu)_+ = r\,\frac{y_i^\star}{\|y^\star\|_2}, \qquad i = 1, \dots, m. \tag{29}$$

Squaring and summing gives

$$\sum_{i=1}^m (g_i - \nu)_+^2 = \frac{r^2}{\|y^\star\|_2^2} \sum_{i=1}^m (y_i^\star)^2 \tag{30}$$

$$= r^2. \tag{31}$$

It remains to identify ν with the optimal objective value. Multiply (28) by $y_i^\star$ and sum over i,

$$g^\top y^\star - \nu \sum_{i=1}^m y_i^\star = \frac{r}{\|y^\star\|_2} \sum_{i=1}^m (y_i^\star)^2.$$

Using $\sum_{i=1}^m y_i^\star = 1$ and $\sum_{i=1}^m (y_i^\star)^2 = \|y^\star\|_2^2$, we obtain $g^\top y^\star - \nu = r\,\|y^\star\|_2$. Therefore

$$\nu = g^\top y^\star - r\,\|y^\star\|_2 = \mathcal{A}(g, h).$$

Writing $\lambda := \mathcal{A}(g, h)$, equation (31) becomes

$$\sum_{i=1}^m (g_i - \lambda)_+^2 = r^2. \tag{32}$$

Thus the optimal value λ is the unique solution of (32). Since the function $s \mapsto \sum_{i=1}^m (g_i - s)_+^2$ is continuous and non-increasing, and since $r > 0$ with probability one, we also obtain, for every s,

$$\mathcal{A}(g, h) > s \quad \Longleftrightarrow \quad \sum_{i=1}^m (g_i - s)_+^2 > \|h\|_2^2\,. \tag{33}$$

Indeed, equality holds at $s = \lambda$; if $s < \lambda$ the left-hand side of (32) is strictly larger than r^2, while if $s > \lambda$ it is strictly smaller.

Step 3: direct integral bounds for the truncated Gaussian moments.

Let $Z \sim N(0,1)$ and define

$$I(s) := \mathbb{E}(Z-s)_+^2, \qquad J(s) := \mathbb{E}(Z-s)_+^4.$$

We derive the estimates needed below directly from the Gaussian integrals, without appealing to Mills-ratio bounds.

By definition,

$$I(s) = \int_s^\infty (z-s)^2 \phi(z)\, dz,$$

where

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Set $u = z - s$, then

$$I(s) = \int_0^\infty u^2 \phi(s+u)\, du = \phi(s) \int_0^\infty u^2 e^{-su-u^2/2}\, du.$$

Now set $w = su$. For $s > 0$,

$$I(s) = \frac{\phi(s)}{s^3} \int_0^\infty w^2 e^{-w-w^2/(2s^2)}\, dw. \tag{34}$$

Moreover,

$$e^{-w-w^2/(2s^2)} \le e^{-w},$$

and therefore

$$I(s) \le \frac{\phi(s)}{s^3} \int_0^\infty w^2 e^{-w}\, dw = 2\,\frac{\phi(s)}{s^3}.$$

For the lower bound, restrict the integral in (34) to $0 \le w \le 1$. Since $s \ge 1$,

$$e^{-w-w^2/(2s^2)} \ge e^{-3/2} \qquad (0 \le w \le 1),$$

and hence

$$I(s) \ge \frac{\phi(s)}{s^3} e^{-3/2} \int_0^1 w^2\, dw = \frac{e^{-3/2}}{3}\,\frac{\phi(s)}{s^3}.$$

Thus, for every $s \ge 1$,

$$\frac{e^{-3/2}}{3}\,\frac{\phi(s)}{s^3} \le I(s) \le 2\,\frac{\phi(s)}{s^3}. \tag{35}$$

Similarly,

$$J(s) = \int_s^\infty (z-s)^4 \phi(z)\, dz.$$

With the same substitutions $u = z - s$ and $w = su$,

$$J(s) = \frac{\phi(s)}{s^5} \int_0^\infty w^4 e^{-w-w^2/(2s^2)}\, dw.$$

For $s \geq 1$,

$$J(s) \leq \frac{\phi(s)}{s^5} \int_0^\infty w^4 e^{-w}\, dw = 24\, \frac{\phi(s)}{s^5}.$$

Hence

$$J(s) \leq 24\, \frac{\phi(s)}{s^5}, \qquad s \geq 1. \tag{36}$$

We apply these estimates at

$$t = (1-\rho)\sqrt{2\log\alpha}.$$

For α sufficiently large, $t \geq 1$. Since

$$e^{-t^2/2} = \alpha^{-(1-\rho)^2},$$

we have

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \alpha^{-(1-\rho)^2}.$$

Therefore, by (35),

$$\alpha I(t) \geq \frac{e^{-3/2}}{3\sqrt{2\pi}} \frac{\alpha^{1-(1-\rho)^2}}{t^3} \tag{37}$$

$$= \frac{e^{-3/2}}{3\sqrt{2\pi}} \frac{\alpha^\beta}{t^3}, \tag{38}$$

where

$$\beta = 1 - (1-\rho)^2 = 2\rho - \rho^2 > 0.$$

Since

$$t^3 = (1-\rho)^3 (2\log\alpha)^{3/2},$$

it follows that

$$\alpha I(t) \to \infty \qquad (\alpha \to \infty).$$

Hence, for sufficiently large α,

$$\alpha I(t) \geq 4. \tag{39}$$

The same direct estimates also give the ratio needed for the second-moment bound:

$$\frac{J(t)}{I(t)^2} \leq \frac{24\, \phi(t) t^{-5}}{\left(\frac{e^{-3/2}}{3} \phi(t) t^{-3}\right)^2}.$$

Thus

$$\frac{J(t)}{I(t)^2} \le 216e^3 \frac{t}{\phi(t)}. \tag{40}$$

Step 4: concentration of the truncated Gaussian sum.

Set

$$X_i := (g_i - t)_+^2, \qquad S_m := \sum_{i=1}^m X_i.$$

Then

$$\mathbb{E}\, S_m = mI(t) = n\alpha I(t) \ge 4n.$$

If $S_m \le 2n$, then

$$|S_m - \mathbb{E}\, S_m| \ge \frac{1}{2}\,\mathbb{E}\, S_m.$$

Chebyshev's inequality along with the independence of the X_i therefore give

$$\mathbb{P}\{S_m \le 2n\} \le 4\,\frac{\mathrm{Var}(S_m)}{(\mathbb{E}\, S_m)^2} = 4\,\frac{m(J(t) - I(t)^2)}{m^2 I(t)^2} \le 4\,\frac{mJ(t)}{m^2 I(t)^2}. \tag{41}$$

Using (40),

$$\frac{J(t)}{I(t)^2} \le 216e^3 \frac{t}{\phi(t)}.$$

we can estimate

$$\mathbb{P}\{S_m \le 2n\} \le C_1 \frac{t}{m\phi(t)} \tag{42}$$

for a constant $C_1 > 0$.

Since $m = n\alpha$ and

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \alpha^{-(1-\rho)^2},$$

we have

$$\frac{t}{m\phi(t)} = \sqrt{2\pi}\,\frac{t}{n\alpha^\beta} \le C_2 \frac{\sqrt{\log \alpha}}{n\alpha^\beta}. \tag{43}$$

On the other hand, since $\|h\|_2^2 \sim \chi_n^2$, a standard chi-square upper-tail bound (See Theorem 2.16 in [3]) gives

$$\mathbb{P}\{\|h\|_2^2 \ge 2n\} \le e^{-cn} \tag{44}$$

for an absolute constant $c > 0$.

By (33),

$$\{\mathcal{A}(g,h) \le t\} = \{S_m \le \|h\|_2^2\} \subseteq \{S_m \le 2n\} \cup \{\|h\|_2^2 \ge 2n\}.$$

Therefore

$$\mathbb{P}\{\mathcal{A}(g,h) \le t\} \le C_2 \frac{\sqrt{\log \alpha}}{n\alpha^\beta} + e^{-cn}. \tag{45}$$

Gordon's inequality (24) now yields

$$\mathbb{P}\{K_{n,m} < t\} \le C\frac{\sqrt{\log \alpha}}{n\alpha^{\beta}} + 2e^{-cn},$$

which proves (21).

Step 5: normalization with an explicit Gaussian norm bound.

Define

$$R_{\max} := \max_{1\le i\le m} \|g_i\|_2 .$$

For $\rho \in (0,1)$, one-sided Gaussian concentration (see Theorem 13.9 and Proposition 13.15 in [3]) gives

$$\mathbb{P}\left\{\|g_i\|_2 \ge (1+\rho)\sqrt{n}\right\} \le e^{-n\rho^2/2}.$$

A union bound gives

$$\mathbb{P}\left\{R_{\max} \ge (1+\rho)\sqrt{n}\right\} \le me^{-n\rho^2/2}. \tag{46}$$

On the event

$$K_{n,m} \ge t > 0$$

we have, for every $v \in S^{n-1}$,

$$\max_i \langle g_i, v\rangle \ge t > 0.$$

Choose $i(v)$ to be the index attaining this maximum. Then

$$\max_i \langle u_i, v\rangle \ge \frac{\max_i \langle g_i, v\rangle}{R_{\max}}.$$

Taking the minimum over v yields

$$\mu_{n,m} \ge \frac{K_{n,m}}{R_{\max}}. \tag{47}$$

Thus, on the intersection of the events $K_{n,m} \ge t$ and $R_{\max} \le (1+\rho)\sqrt{n}$, we obtain

$$\mu_{n,m} \ge \frac{1-\rho}{1+\rho}\sqrt{\frac{2\log\alpha}{n}}. \tag{48}$$

For $0 < \rho < 1$,

$$\frac{1-\rho}{1+\rho} = 1 - \frac{2\rho}{1+\rho} \ge 1 - 2\rho.$$

Since $\rho = \varepsilon/3$,

$$1 - 2\rho = 1 - \frac{2\varepsilon}{3} \ge 1 - \varepsilon.$$

Hence

$$\mu_{n,m} \geq (1-\varepsilon)\sqrt{\frac{2\log\alpha}{n}}$$

on this intersection.

Combining (21) and (46), and using $\rho = \frac{\varepsilon}{3}$, gives

$$\mathbb{P}\left\{\mu_{n,m} < (1-\varepsilon)\sqrt{\frac{2\log\alpha}{n}}\right\} \leq C\frac{\sqrt{\log\alpha}}{n\alpha^{\beta}} + 2e^{-cn} + m\exp\left(-\frac{n\varepsilon^2}{18}\right). \tag{49}$$

Since $\alpha = m/n$, this is exactly (22).

Finally, if

$$\alpha = \frac{m}{n} \to \infty, \qquad \log\alpha = o(n),$$

then

$$\frac{\sqrt{\log\alpha}}{n\alpha^{\beta}} \to 0, \qquad e^{-cn} \to 0,$$

and

$$me^{-n\varepsilon^2/18} = \exp\left\{\log m - \frac{n\varepsilon^2}{18}\right\} \to 0,$$

because

$$\log m = \log n + \log\alpha = o(n).$$

The proof is complete. □

5.3.2 Upper bound for the Gaussian cosine measure

Theorem 5.4. *There exist absolute constants $C_* > 0$ and $\alpha_0 > 1$ such that, for every $\alpha \geq \alpha_0$, with*

$$K := \sqrt{2\log\alpha},$$

one has

$$\mathbb{P}\{K_{n,m} > K\} \leq 1 - \sqrt{1 - \frac{2C_*}{\sqrt{2\pi}\,K}}. \tag{50}$$

Consequently, for every fixed $\varepsilon > 0$ and all sufficiently large α,

$$\mathbb{P}\left\{\mu_{n,m} > (1+\varepsilon)\sqrt{\frac{2\log(m/n)}{n}}\right\} \leq 1 - \sqrt{1 - \frac{2C_*}{\sqrt{2\pi}\sqrt{2\log(m/n)}}} + m\exp\left\{-\frac{n}{2}\left(\frac{\varepsilon}{1+\varepsilon}\right)^2\right\}. \tag{51}$$

In particular, if

$$\frac{m}{n} \to \infty, \qquad \log \frac{m}{n} = o(n),$$

then, for every fixed $\varepsilon > 0$*,*

$$\mathbb{P}\left\{ \mu_{n,m} > (1+\varepsilon)\sqrt{\frac{2\log(m/n)}{n}} \right\} \longrightarrow 0.$$

The proof combines three standard ingredients: a weighted second-moment construction for perceptron feasibility, following ideas used for the negative spherical perceptron; exponential centering of the one-dimensional weight (these ideas are all inspired by [27]); and the Hermite-rank/Mehler representation of correlations of functions of Gaussian variables [6, 23]. The specific observation exploited here is that centering annihilates the first Hermite coefficient. Combined with the exact moments of spherical overlaps, this gives a direct second-moment bound without optimizing a nontrivial overlap functional.

Proof. **Step 1: a centered exponential weight.**
Set

$$K := \sqrt{2\log \alpha}.$$

For $c > 0$, define

$$f_c(x) := e^{-cx}\mathbf{1}_{\{x \geq -K\}}, \qquad p(c) := \mathbb{E}\, f_c(G),$$

where $G \sim N(0,1)$.

Writing the expectation as an integral and completing the square,

$$p(c) = \frac{1}{\sqrt{2\pi}} \int_{-K}^{\infty} e^{-cx - x^2/2}\, dx \tag{52}$$

$$= e^{c^2/2} \frac{1}{\sqrt{2\pi}} \int_{-K}^{\infty} e^{-(x+c)^2/2}\, dx \tag{53}$$

$$= e^{c^2/2}\Phi(K-c). \tag{54}$$

Noting that for $c \in [0, C]$ the difference quotients of $f_c(G)$ with respect to c are dominated by $e^{CK}|G|$, dominated convergence gives

$$p'(c) = -\,\mathbb{E}\left[G e^{-cG} \mathbf{1}_{\{G \geq -K\}} \right],$$

Thus we have

$$p'(0) = -\phi(K) < 0,$$

while

$$p(c) \geq e^{cK/2}\mathbb{P}\{-K \leq G \leq -K/2\} \to \infty \text{ as } c \to \infty,$$

guaranteeing the existence of a $c = c_K > 0$ with

$$p'(c) = 0.$$

This is equivalent to

$$\mathbb{E}[Gf_c(G)] = 0. \tag{55}$$

Differentiating (54) gives the critical-point equation

$$c\Phi(K - c) = \phi(K - c), \tag{56}$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Put

$$t = K - c.$$

Then

$$K = t + \frac{\phi(t)}{\Phi(t)}.$$

Let

$$f\colon\ t \longmapsto t + \frac{\phi(t)}{\Phi(t)}$$

and note that $f'(t) = \mathrm{Var}(Z|Z \le t) > 0$ where $Z \sim N(0,1)$. So f is continuous and strictly increasing with

$$\lim_{t\to-\infty} f(t) = 0 \text{ and } \lim_{t\to\infty} f(t) = \infty.$$

Hence, for every $K > 0$, there is a unique t, and therefore a unique $c = K - t > 0$.

As $K \to \infty$, necessarily $t \to \infty$. From

$$c = \frac{\phi(t)}{\Phi(t)}$$

we obtain

$$c \to 0$$

and

$$tc \le 2t\phi(t) \to 0$$

for all sufficiently large t, since $\Phi(t) \ge 1/2$. Consequently

$$Kc = tc + c^2 \longrightarrow 0. \tag{57}$$

It will be convenient in our quantitative estimates below to use the following bound: There exists K_0 such that, for all $K \ge K_0$,

$$Kc \le \frac{1}{8e^2}. \tag{58}$$

Define

$$f := f_{c_K}, \qquad p := \mathbb{E}\, f(G), \qquad F := \frac{f}{p}, \qquad V := \mathrm{Var}(F(G)).$$

Then

$$\mathbb{E}\, F(G) = 1, \qquad \mathbb{E}[GF(G)] = 0, \tag{59}$$

where the second equality follows from (55).

Step 2: an explicit variance bound from direct Gaussian integrals.

Let

$$r := \Phi(-K) = \int_K^\infty \phi(x)\, dx.$$

With $x = K + u$,

$$r = \phi(K) \int_0^\infty e^{-Ku - u^2/2}\, du.$$

Setting $w = Ku$ gives

$$r = \frac{\phi(K)}{K} \int_0^\infty e^{-w - w^2/(2K^2)}\, dw. \tag{60}$$

For $K \geq 1$, restricting to $0 \leq w \leq 1$ yields

$$r \geq e^{-3/2} \frac{\phi(K)}{K}. \tag{61}$$

For $a = 1, 2$, define the shifted tails

$$r_a := 1 - \Phi(K - ac) = \int_{K-ac}^\infty \phi(x)\, dx.$$

Since $K - ac < K$,

$$r_a - r = \int_{K-ac}^K \phi(x)\, dx.$$

For sufficiently large K, the Gaussian density is decreasing throughout this interval, so

$$0 \leq r_a - r \leq ac\, \phi(K - ac). \tag{62}$$

Furthermore,

$$\frac{\phi(K - ac)}{\phi(K)} = \exp\left(aKc - \frac{a^2c^2}{2}\right) \leq e^{aKc}. \tag{63}$$

Using (58), for $a = 1, 2$,

$$\phi(K - ac) \leq e^{1/(4e^2)} \phi(K) < e^{1/2} \phi(K).$$

Combining this with (61) and using (58) gives

$$\begin{aligned}\frac{r_a - r}{r} &\le ace^{1/2}\phi(K)\frac{Ke^{3/2}}{\phi(K)} \\ &= ae^2Kc \\ &\le \frac{a}{8}.\end{aligned}$$

Hence by (62) and the above we obtain

$$r \le r_1 \le \frac{9}{8}r, \qquad r \le r_2 \le \frac{5}{4}r. \tag{64}$$

We also need an explicit bound on c^2 relative to r. From (56),

$$c = \frac{\phi(K-c)}{\Phi(K-c)}.$$

For large K, $K - c > 0$, and hence $\Phi(K-c) \ge 1/2$. So by (63)

$$c \le 2\phi(K-c) \le 2e^{Kc}\phi(K).$$

Using (58),

$$c \le 3\phi(K)$$

for all sufficiently large K. Thus, by (61),

$$\frac{c^2}{r} \le 9e^{3/2}K\phi(K).$$

Since $K\phi(K) \to 0$, after increasing K_0 if necessary we may assume

$$c^2 \le \frac{1}{8}r. \tag{65}$$

Now

$$p = \mathbb{E}\, f(G) = e^{c^2/2}\Phi(K-c).$$

Recall that

$$f(x) = e^{-cx}\mathbf{1}_{\{x \ge -K\}},$$

so

$$f(x)^2 = e^{-2cx}\mathbf{1}_{\{x \ge -K\}},$$

Thus

$$\begin{aligned}\mathbb{E}[f(G)^2] &= \frac{1}{\sqrt{2\pi}}\int_{-K}^{\infty} e^{-2cx - x^2/2}\,dx. \\ &= e^{2c^2}\frac{1}{\sqrt{2\pi}}\int_{-K+2c}^{\infty} e^{-u^2/2}\,du \\ &= e^{2c^2}\left[1 - \Phi(-K+2c)\right] \\ &= e^{2c^2}\Phi(K-2c).\end{aligned}$$

Noting that

$$V = \mathbb{E}[F(G)^2] - \big(\mathbb{E}[F(G)]\big)^2 = \frac{\mathbb{E}[f(G)^2]}{p^2} - 1,$$

we have

$$1 + V = e^{c^2} \frac{\Phi(K - 2c)}{\Phi(K - c)^2} = e^{c^2} \frac{1 - r_2}{(1 - r_1)^2}. \tag{66}$$

For large K we may also assume $r \le 1/8$. Then

$$r_1 \le \frac{9}{8} r < \frac{1}{4}.$$

Also for $0 \le x \le 1/4$,

$$(1 - x)^{-2} \le 1 + 4x.$$

Moreover, since $c^2 \le r/8$,

$$e^{c^2} \le 1 + 2c^2 \le 1 + \frac{r}{4}.$$

Thus, using $1 - r_2 \le 1$ in (66),

$$\begin{aligned} 1 + V &\le \left(1 + \frac{r}{4}\right)(1 + 4r_1) \\ &\le \left(1 + \frac{r}{4}\right)\left(1 + \frac{9}{2} r\right). \end{aligned}$$

For $r \le 1/8$, the right-hand side is at most $1 + 6r$. Therefore

$$V \le C_* \, \Phi(-K). \tag{67}$$

For example, we can choose $C_* = 6$.

From (60)

$$\Phi(-K) \le \frac{\phi(K)}{K}.$$

Since $K = \sqrt{2 \log \alpha}$ and $e^{-K^2/2} = \alpha^{-1}$, we obtain

$$\alpha V \le \frac{C_*}{\sqrt{2\pi}\, K}. \tag{68}$$

Step 3: centering gives a global quadratic correlation bound.

Let (G_1, G_2) be jointly standard Gaussian with correlation $q \in (-1, 1)$. We use the Hermite polynomials in the "probabilist's normalization", namely

$$H_j(x) = (-1)^j e^{x^2/2} \frac{d^j}{dx^j} e^{-x^2/2},$$

which satisfy

$$\mathbb{E}[H_j(G) H_k(G)] = j! \, \mathbf{1}_{\{j=k\}}.$$

Expand

$$F(x) = \sum_{j=0}^{\infty} \frac{a_j}{j!} H_j(x).$$

By (59), we have $a_0 = 1$ and $a_1 = 0$.

Recall Mehler's identity ([29] 18.18.28) for this normalization (see 18.7.11 there)

$$\frac{1}{\sqrt{1-q^2}} \exp\left(\frac{qxy - \frac{1}{2}q^2(x^2+y^2)}{1-q^2} \right) = \sum_{j=0}^{\infty} \frac{q^j}{j!} H_j(x) H_j(y).$$

Multiplying each side by the product of two independent standard Gaussian pdfs gives the following formula for the joint distribution

$$\phi_q(x,y) = \phi(x)\phi(y) \sum_{j=0}^{\infty} \frac{q^j}{j!} H_j(x) H_j(y).$$

This gives us the following identity, since F is square-integrable with respect to the standard Gaussian measure:

$$\begin{aligned}
\mathbb{E}[F(G_1)F(G_2)] &= \int_{\mathbb{R}^2} F(x)F(y)\phi_q(x,y)\,dx\,dy \\
&= \sum_{j=0}^{\infty} \frac{q^j}{j!} \left(\int_{\mathbb{R}} F(x) H_j(x) \phi(x)\,dx \right) \left(\int_{\mathbb{R}} F(y) H_j(y) \phi(y)\,dy \right) \\
&= 1 + \sum_{j=2}^{\infty} \frac{a_j^2}{j!} q^j
\end{aligned}$$

Since, for $j \geq 2$ and $|q| \leq 1$,

$$q^j \leq |q|^j \leq q^2,$$

we obtain

$$\mathbb{E}[F(G_1)F(G_2)] \leq 1 + q^2 \sum_{j=2}^{\infty} \frac{a_j^2}{j!}. \tag{69}$$

By Parseval's identity,

$$\sum_{j=2}^{\infty} \frac{a_j^2}{j!} = \mathrm{Var}(F(G)) = V.$$

Hence

$$\mathbb{E}[F(G_1)F(G_2)] \leq 1 + Vq^2. \tag{70}$$

The same bound holds at $q = \pm 1$ by Cauchy–Schwarz.

Step 4: second moment of the weighted feasible volume.

Let σ denote normalized surface measure on S^{n-1} and set

$$Z := \int_{S^{n-1}} \prod_{i=1}^{m} f(\langle g_i, v\rangle)\, d\sigma(v).$$

Since $f(x) = 0$ when $x < -K$, the event $Z > 0$ implies the existence of $v \in S^{n-1}$ such that

$$\langle g_i, v\rangle \geq -K, \qquad i = 1, \ldots, m. \tag{71}$$

For each fixed v, the random variables

$$\langle g_1, v\rangle, \ldots, \langle g_m, v\rangle$$

are independent standard Gaussians, as they are linear combinations of such Gaussians. Hence

$$\mathbb{E}\, Z = p^m.$$

Let V_1, V_2 be independent uniform random vectors on S^{n-1} and set

$$Q := \langle V_1, V_2\rangle.$$

Conditioned on $Q = q$, each pair

$$\big(\langle g_i, V_1\rangle, \langle g_i, V_2\rangle\big)$$

is jointly standard Gaussian with correlation q. Therefore

$$\begin{aligned}
\frac{\mathbb{E}\, Z^2}{(\mathbb{E}\, Z)^2} &= \frac{\mathbb{E}_Q\left[\mathbb{E}[f(G_1)f(G_2)]^m\right]}{p^{2m}} \\
&= \mathbb{E}_Q\left[\mathbb{E}[F(G_1)F(G_2)]^m\right] \\
&\leq \mathbb{E}_Q(1 + VQ^2)^m \\
&\leq \mathbb{E}_Q\, e^{mVQ^2},
\end{aligned}$$

where in the last two lines we used (70) and the inequality $1 + x \leq e^x$.

We need the following lemma in the next calculation. Its proof can be found in Appendix A.

Lemma 5.5. *The even moments of Q are*

$$\mathbb{E}\, Q^{2j} = \frac{(2j-1)!!}{n(n+2)\cdots(n+2j-2)} \leq \frac{(2j-1)!!}{n^j}.$$

By this lemma, whenever $2mV < n$,

$$\begin{aligned}\mathbb{E}\, e^{mVQ^2} &\le \sum_{j=0}^{\infty} \frac{(mV)^j}{j!} \frac{(2j-1)!!}{n^j} \\ &= \left(1 - \frac{2mV}{n}\right)^{-1/2} \\ &= (1 - 2\alpha V)^{-1/2},\end{aligned}$$

where we used the binomial series expansion for $(1-x)^{-1/2}$.

Using (68), for sufficiently large α we have

$$2\alpha V < 1,$$

and

$$\frac{\mathbb{E}\, Z^2}{(\mathbb{E}\, Z)^2} \le \left(1 - \frac{2C_*}{\sqrt{2\pi}\, K}\right)^{-1/2}. \tag{72}$$

The Paley–Zygmund inequality gives

$$\mathbb{P}(Z > 0) \ge \frac{(\mathbb{E}\, Z)^2}{\mathbb{E}\, Z^2},$$

and therefore

$$\mathbb{P}(Z > 0) \ge \sqrt{1 - \frac{2C_*}{\sqrt{2\pi}\, K}}. \tag{73}$$

On the event $Z > 0$, choose v satisfying (71). Replacing v by $-v$ gives

$$\langle g_i, -v \rangle \le K \qquad \text{for all } i.$$

Hence

$$K_{n,m} \le K.$$

Consequently

$$\mathbb{P}\{K_{n,m} > K\} \le 1 - \sqrt{1 - \frac{2C_*}{\sqrt{2\pi}\, K}},$$

which proves (50).

Step 5: normalization with an explicit Gaussian norm tail.

Fix $\eta \in (0,1)$ and define

$$R_{\min} := \min_{1 \le i \le m} \|g_i\|_2 .$$

The standard Gaussian norm lower-tail estimate gives

$$\mathbb{P}\left\{\|g_i\|_2 \le (1-\eta)\sqrt{n}\right\} \le e^{-n\eta^2/2}.$$

By the union bound,

$$\mathbb{P}\left\{R_{\min} \le (1-\eta)\sqrt{n}\right\} \le me^{-n\eta^2/2}. \tag{74}$$

On the event

$$K_{n,m} \le K, \qquad R_{\min} > (1-\eta)\sqrt{n},$$

there exists $v \in S^{n-1}$ such that

$$\langle g_i, v\rangle \le K \qquad \text{for every } i.$$

Then

$$\langle u_i, v\rangle = \frac{\langle g_i, v\rangle}{\|g_i\|_2} \le \frac{K}{\|g_i\|_2} \le \frac{K}{(1-\eta)\sqrt{n}}.$$

Hence

$$\mu_{n,m} \le \frac{1}{1-\eta}\sqrt{\frac{2\log\alpha}{n}}.$$

Combining (50) and (74) and union bounding gives,

$$\begin{aligned}
&\mathbb{P}\left\{\mu_{n,m} > \frac{1}{1-\eta}\sqrt{\frac{2\log\alpha}{n}}\right\} \\
&\qquad \le \mathbb{P}\{K_{n,m} > K\} + \mathbb{P}\{R_{\min} \le (1-\eta)\sqrt{n}\} \\
&\qquad \le 1 - \sqrt{1 - \frac{2C_*}{\sqrt{2\pi}\sqrt{2\log\alpha}}} + me^{-n\eta^2/2}.
\end{aligned}$$

Given $\varepsilon > 0$, choose

$$\eta = \frac{\varepsilon}{1+\varepsilon}.$$

Then

$$\frac{1}{1-\eta} = 1 + \varepsilon,$$

which gives (51).

Finally, if

$$\alpha = \frac{m}{n} \to \infty, \qquad \log\alpha = o(n),$$

then the first term on the right-hand side of (51) tends to zero because $\sqrt{\log\alpha} \to \infty$, while

$$me^{-n\eta^2/2} = \exp\left\{\log m - \frac{n\eta^2}{2}\right\} \to 0$$

because $\log m = \log n + \log\alpha = o(n)$.

□

Acknowledgements

T.S. wants to thank Benjamin Friedlander, whose inquisitive questions about conic combinations inspired the research presented in this paper, and March Boedihardjo for helpful discussions on conic combinations. The authors acknowledge support from NIH R01HL163515, NSF DMS-2208356, and DOE DE-SC0023490.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, Nov. 2006. doi: 10.1109/TSP.2006.881199.

[2] B. Alexeev, J. Cahill, and D. G. Mixon. Full spark frames. *Journal of Fourier Analysis and Applications*, 18(6):1167–1194, 2012. doi: 10.1007/s00041-012-9235-4.

[3] A. Bandeira, A. Singer, and T. Strohmer. Mathematics of data science. arXiv:2607.11938, 2026.

[4] G. Bonnet and E. O'Reilly. Facets of spherical random polytopes. *Mathematische Nachrichten*, 295(10):1901–1933, 2022. doi: 10.1002/mana.202000314.

[5] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[6] P. Breuer and P. Major. Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13:425–441, 1983. doi: 10.1016/0047-259X(83)90019-2.

[7] A. Brøndsted. *An introduction to convex polytopes*. Number 90 in Graduate texts in mathematics. Springer, New York Heidelberg, 1983. ISBN 978-0-387-90722-2.

[8] P. G. Casazza and G. Kutyniok. *Finite frames: Theory and applications*. Birkhäuser Boston, MA, 2012. ISBN 978-0-8176-8372-6.

[9] O. Christensen. *Frames and bases: An introductory course*. Birkhäuser Boston, MA, 2008. ISBN 978-0-8176-4677-6.

[10] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. SIAM, 2009. doi: 10.1137/1.9780898718768.

[11] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*, volume 290 of *Grundlehren der mathematischen Wissenschaften*. Springer New York, New York, NY, 1999. ISBN 978-1-4419-3134-4. doi: 10.1007/978-1-4757-6568-7.

[12] I. Daubechies. *Ten lectures on wavelets*. SIAM, 1992. ISBN 978-0-89871-274-2.

[13] C. Davis. Theory of positive linear dependence. *American Journal of Mathematics*, 76(4):733–746, 1954. doi: 10.2307/2372648.

[14] M. Dodangeh, L. N. Vicente, and Z. Zhang. On the optimal order of worst case complexity of direct search. *Optimization Letters*, 10:699–708, 2016. doi: 10.1007/s11590-015-0908-1.

[15] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003. doi: 10.1073/pnas.0437847100.

[16] M. Dostert, D. de Laat, and P. Moustrou. Exact semidefinite programming bounds for packing problems, 2020. URL `https://arxiv.org/abs/2001.00256`.

[17] B. Friedlander. Personal Communications with T. Strohmer, 2023, 2024.

[18] R. H. Hardin, N. J. A. Sloane, and W. D. Smith. Spherical coverings. `http://neilsloane.com/coverings/`, 1994. Tables of putatively optimal coverings of the sphere by n equal caps; now maintained at `https://cohn.mit.edu/sloane`.

[19] W. Hare and G. Jarry-Bolduc. A deterministic algorithm to compute the cosine measure of a finite positive spanning set. *arXiv preprint*, 2020. arXiv:2004.03507.

[20] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Grundlehren text editions. Springer Berlin, Heidelberg, 2001. ISBN 978-3-540-42205-1.

[21] P. O. Hoyer. Non-negative sparse coding, 2002. arXiv:cs/0202009.

[22] R. Jagannathan and T. Ma. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *The Journal of Finance*, 58(4):1651–1683, Aug. 2003. doi: 10.1111/1540-6261.00580.

[23] S. Janson. *Gaussian Hilbert Spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997. ISBN 9780511526169.

[24] J. Larson, M. Menickelly, and S. M. Wild. Derivative-free optimization methods. *Acta Numerica*, 28:287–404, 2019. doi: 10.1017/s0962492919000060.

[25] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999. doi: 10.1038/44565.

[26] R. L. McKinney. Positive bases for linear spaces. *Transactions of the American Mathematical Society*, 103(1):131–148, 1962. doi: 10.2307/1993744.

[27] A. Montanari, Y. Zhong, and K. Zhou. Tractability from overparametrization: the example of the negative perceptron. *Probability Theory and Related Fields*, 188:805–910, 2024. doi: 10.1007/s00440-023-01248-y.

[28] G. Naevdal. Positive bases with maximal cosine measure. *Optimization Letters*, 13(6):1381–1388, 2019. doi: 10.1007/s11590-018-1334-y.

[29] F. W. J. Olver et al. NIST Digital Library of Mathematical Functions. `https://dlmf.nist.gov/`.

[30] V. K. Potluru, S. M. Plis, S. Luan, V. D. Calhoun, and T. P. Hayes. Sparseness and a reduction from Totally Nonnegative Least Squares to SVM. In *The 2011 International Joint Conference on Neural Networks*, pages 1922–1929, San Jose, CA, USA, 2011. IEEE. doi: 10.1109/IJCNN.2011.6033459.

[31] R. G. Regis. On the properties of positive spanning sets and positive bases. *Optimization and Engineering*, 17(1):229–262, 2016. doi: 10.1007/s11081-015-9286-x.

[32] P. A. Samuelson. *Foundations of Economic Analysis*. Harvard University Press, Cambridge, MA, enlarged ed., 1983 edition, 1983. ISBN 9780674313033.

[33] P. Smaragdis, B. Raj, and M. Shashanka. A probabilistic latent variable model for acoustic modeling. In *NIPS Workshop on Advances in Models for Acoustic Processing*, Whistler, BC, Canada, 2006.

[34] T. Strohmer and R. Heath Jr. Grassmannian frames with applications to coding and communications. *Appl. Comp. Harm. Anal.*, 14(3):257–275, 2003. doi: 10.1016/S1063-5203(03)00023-X.

[35] C. Thrampoulidis, S. Oymak, and B. Hassibi. The Gaussian min–max theorem in the presence of convexity. arXiv:1408.4837, 2014.

[36] B. D. Van Veen and K. M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE assp magazine*, 5(2):4–24, 1988. doi: 10.1109/53.665.

[37] T. Virtanen. Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3):1066–1074, Mar. 2007. doi: 10.1109/TASL.2006.885253.

[38] T. Virtanen, J. F. Gemmeke, B. Raj, and P. Smaragdis. Compositional Models for Audio Processing: Uncovering the structure of sound mixtures. *IEEE Signal Processing Magazine*, 32(2):125–144, Mar. 2015. doi: 10.1109/MSP.2013.2288990.

[39] S. F. D. Waldron. *An introduction to finite tight frames*. Birkhäuser New York, NY, 2018. ISBN 978-0-8176-4814-5.

[40] L. Welch. Lower bounds on the maximum cross correlation of signals (Corresp.). *IEEE Transactions on Information Theory*, 20(3):397–399, May 1974. doi: 10.1109/TIT.1974.1055219.

[41] J. G. Wendel. A problem in geometric probability. *Mathematica Scandinavica*, 11(1):109–111, 1962. doi: 10.7146/math.scand.a-10655.

[42] R. M. Young. *An introduction to nonharmonic Fourier series*, volume 93. Academic press, 1981. ISBN 978-0-12-772850-6.

[43] G. M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer New York, 1995. ISBN 978-0-387-94365-7.

A. Proof of Lemma 5.5

Lemma 5.5. *The even moments of Q are*

$$\mathbb{E}\, Q^{2j} = \frac{(2j-1)!!}{n(n+2)\cdots(n+2j-2)} \le \frac{(2j-1)!!}{n^j}.$$

Proof. First, reduce the inner product to one coordinate. Condition on $V_1 = v$, and choose an orthogonal matrix O sending v to e_1. Then

$$\langle v, V_2\rangle = \langle Ov, OV_2\rangle = \langle e_1, OV_2\rangle.$$

Since OV_2 is still uniform on the sphere, this has the same distribution as the first coordinate U_1 of a uniform vector U. This holds for every v, so

$$Q \overset{d}{=} U_1.$$

Next, express U using a Gaussian vector. Let

$$X = (X_1, \dots, X_n) \sim N(0, I_n), \qquad R = \|X\|, \qquad U = \frac{X}{R}.$$

Then U is uniform on the sphere and independent of R. These facts follow from the Gaussian density in polar coordinates: with respect to $dr\, d\sigma(u)$, it is

$$C_n r^{n-1} e^{-r^2/2},$$

which factors into a density depending only on r and uniform surface measure in u. Because $X_1 = RU_1$, independence gives

$$\mathbb{E}X_1^{2j} = \mathbb{E}[R^{2j}U_1^{2j}] = \mathbb{E}R^{2j}\,\mathbb{E}U_1^{2j}.$$

Consequently,

$$\mathbb{E}Q^{2j} = \mathbb{E}U_1^{2j} = \frac{\mathbb{E}X_1^{2j}}{\mathbb{E}R^{2j}}. \tag{75}$$

We can calculate both moments by integration by parts. For the numerator, use $\phi'(x) = -x\phi(x)$:

$$\begin{aligned}\mathbb{E}X_1^{2j} &= \int_{\mathbb{R}} x^{2j}\phi(x)\,dx \\ &= -\int_{\mathbb{R}} x^{2j-1}\phi'(x)\,dx \\ &= (2j-1)\int_{\mathbb{R}} x^{2j-2}\phi(x)\,dx \\ &= (2j-1)\mathbb{E}X_1^{2j-2}.\end{aligned}$$

The boundary term vanishes because the Gaussian density decays faster than any polynomial. Starting from $\mathbb{E}X_1^0 = 1$, this recurrence gives

$$\mathbb{E}X_1^{2j} = (2j-1)!!.$$

For the denominator, the radial density above gives

$$\mathbb{E}R^{2j} = C_n \int_0^\infty r^{n+2j-1}e^{-r^2/2}\,dr.$$

Write the integrand as $r^{n+2j-2}(re^{-r^2/2})$, and use

$$\frac{d}{dr}e^{-r^2/2} = -re^{-r^2/2}.$$

Integration by parts yields

$$\begin{aligned}\mathbb{E}R^{2j} &= (n+2j-2)C_n \int_0^\infty r^{n+2j-3}e^{-r^2/2}\,dr \\ &= (n+2j-2)\mathbb{E}R^{2j-2}.\end{aligned}$$

Again the boundary terms vanish. Starting from $\mathbb{E}R^0 = 1$, we obtain

$$\mathbb{E}R^{2j} = n(n+2)\cdots(n+2j-2).$$

Substituting into (75) proves the formula. Finally, every denominator factor is at least n, so

$$\mathbb{E}Q^{2j} \le \frac{(2j-1)!!}{n^j}.$$

For $j = 0$, both sides are 1, using the conventions that an empty product is 1 and $(-1)!! = 1$. □

B. Explanation of the inaccuracy in Bonnet and O'Reilly

Using the notation in [4, Lemma 6.10], with n points in dimension d, Bonnet and O'Reilly claim there exist constants C_1, C_2 such that:

$$C_1 \frac{r_1 d \ln(n/d)}{n} \;\geq\; C_2 \frac{r_1 d \ln n}{n-d}.$$

We now demonstrate that for a certain choice of n this does not hold. Choose

$$n = \lceil d \log d \rceil, \qquad d \to \infty.$$

The ceiling operation introduces an error smaller than 1, so we can write

$$n = d \log d + \delta_d, \qquad 0 \leq \delta_d < 1.$$

Consequently,

$$\frac{n}{d} = \log d + \frac{\delta_d}{d}.$$

The basic limit used below is

$$\frac{\log \log d}{\log d} \longrightarrow 0.$$

Indeed, setting $x = \log d$, this becomes the familiar limit $\log x / x \to 0$. 1. The first expression satisfies

$$\frac{d \log(n/d)}{n} \sim \frac{\log \log d}{\log d} \longrightarrow 0.$$

First divide its numerator and denominator by d:

$$\frac{d \log(n/d)}{n} = \frac{\log(n/d)}{n/d}.$$

We already know that

$$\frac{n}{d} = \log d + \frac{\delta_d}{d} \sim \log d.$$

For the numerator, factor $\log d$ inside the logarithm:

$$\begin{aligned} \log(n/d) &= \log \left(\log d + \frac{\delta_d}{d} \right) \\ &= \log \left[\log d \left(1 + \frac{\delta_d}{d \log d} \right) \right] \\ &= \log \log d + \log \left(1 + \frac{\delta_d}{d \log d} \right). \end{aligned}$$

The final term tends to zero because

$$0 \leq \frac{\delta_d}{d \log d} < \frac{1}{d \log d} \longrightarrow 0.$$

Thus

$$\log(n/d) = \log\log d + o(1) \sim \log\log d.$$

Combining the numerator and denominator estimates gives

$$\frac{\log(n/d)}{n/d} \sim \frac{\log\log d}{\log d} \longrightarrow 0.$$

Here $f_d \sim g_d$ means $f_d/g_d \to 1$. 2. The second expression satisfies

$$\frac{d\log n}{n-d} \longrightarrow 1.$$

Again divide the numerator and denominator by d:

$$\frac{d\log n}{n-d} = \frac{\log n}{n/d - 1}.$$

For the numerator, use $n = d(n/d)$:

$$\begin{aligned}\log n &= \log d + \log(n/d) \\ &= \log d + \log\log d + o(1).\end{aligned}$$

The denominator is

$$\frac{n}{d} - 1 = \log d - 1 + \frac{\delta_d}{d}$$

Therefore,

$$\frac{d\log n}{n-d} = \frac{\log d + \log\log d + o(1)}{\log d - 1 + \delta_d/d}. = \frac{1 + \dfrac{\log\log d}{\log d} + \dfrac{o(1)}{\log d}}{1 - \dfrac{1}{\log d} + \dfrac{\delta_d}{d\log d}} \to 1.$$

Thus the desired constants cannot exist.