Repeated Minimizers of $p$-Frame Energies

Alexey Glazyrin
The University of Texas Rio Grande Valley

Josiah Park

Follow this and additional works at: https://scholarworks.utrgv.edu/mss_fac

Recommended Citation
Glazyrin, A., & Park, J. (2020). Repeated Minimizers of $p$-Frame Energies. SIAM Journal on Discrete Mathematics, 34(4), 2411–2423. https://doi.org/10.1137/19M1282702
REPEATED MINIMIZERS OF $p$-FRAME ENERGIES

ALEXEY GLAZYRIN AND JOSIAH PARK

Abstract. For a collection of $N$ unit vectors $X = \{x_i\}_{i=1}^N$, define the $p$-frame energy of $X$ as the quantity $\sum_{i \neq j} |\langle x_i, x_j \rangle|^p$. In this paper, we connect the problem of minimizing this value to another optimization problem, so giving new lower bounds for such energies. In particular, for $p < 2$, we prove that this energy is at least $2(N - d)p^{-\frac{p}{2}}(2 - p)^{\frac{p}{2}} - 2$ which is sharp for $d \leq N \leq 2d$ and $p = 1$. We prove that for $1 \leq m < d$, a repeated orthonormal basis construction of $N = d + m$ vectors minimizes the energy over an interval, $p \in [1, p_m]$, and demonstrate an analogous result for all $N$ in the case $d = 2$. Finally, in connection, we give conjectures on these and other energies.

1. Introduction

Let $A = A_{i,j}$ be an $N \times N$ real matrix of rank less than or equal to $d$, and with ones along the diagonal. The $p$-frame energy of matrix $A$ is denoted

$$E_p(A) = \sum_{i \neq j} |A_{ij}|^p.$$ 

An interesting question is what the optimizing matrices for $E_p(A)$ are for fixed $p$, $N$, and $d$. Bukh and Cox in [3] recently studied the question of bounding $E_\infty(A) = \max\{|A_{ij}|\}$ and its consequences. One special case of this problem concerns matrices associated with unit vectors, $X = \{x_i\}_{i=1}^N \subset \mathbb{R}^d$, in which case $A$ is the Gram matrix of $X$ and so is additionally symmetric and positive semi-definite.

When $p$ is an even natural number, and $N$ is sufficiently large, sharp bounds for energies in the real and complex case follow from the work of Sidelnikov and Welch [20, 22]. Several related results and conjectures on minimizers of $p$-frame energies were formulated by Ehler and Okoudjou in [7]. Using the fact that for $d = N$, the minimizer of $E_p(A)$ for $p = 2$ is an orthonormal basis, they show that whenever $N$ is divisible by $d$ and $p \in (0, 2)$, a repeated orthonormal basis is the unique minimizer. Since the value of the $p$-frame energy does not change when substituting a vector by its opposite, uniqueness in this scenario is meant up to symmetry. Alternatively, uniqueness can be understood by considering the energy on projective spaces $\mathbb{RP}^{d-1}$ (as is done in [2]), where the points in $\mathbb{RP}^{d-1}$ may be identified with lines through the origin in $\mathbb{R}^d$.

The problem of minimizing the 1-frame energy was also posed in [16], where it was conjectured that for any $N$, the repeated orthonormal basis is the unique minimizer. In 1959, Fejes Tóth posed the question [8]: what is the largest sum of non-obtuse angles formed by $N$ lines in $\mathbb{R}^d$? The conjecture stands that for any $N$ the maximum is uniquely attained on a collection of $d$ lines generated by a repeated orthonormal basis and is resolved only for $d = 2$, and for very few cases of $N$ for $d = 3$. The asymptotic result for $d = 3$ is wide open (see [1] and [9] for the recent progress). We note that the conjecture about $E_1$ from [16] immediately follows from the conjecture of Fejes Tóth since $\arccos t \geq \frac{\pi}{2}(1 - t)$ for $t \in [0, 1]$ with equality holding precisely at $t = 0$ and $t = 1$. 

Date: August 13, 2019.
In this paper, we develop new methods for finding lower bounds on $E_p(A)$, based on the framework of Bukh and Cox from [3]. Doing so allows us to prove new general bounds for $E_p(A)$ when $p < 2$. Such bounds are sharp in some cases, particularly, for $p = 1$ and $N \in [d + 1, 2d]$. We also give sharp bounds for $N = d + m$ and $p \in [1, 2\log\frac{2m+1}{2m}/\log\frac{m+1}{m}]$, thus, in particular, partially confirming a conjecture from [5].

Although everything in Sections 2-3 is formulated in the real case, all observations and proofs there work for matrices of complex numbers or quaternions without any changes. Our methods work for general matrix optimization problems, so most of our results will be formulated for matrices. However, we slightly abuse terminology when talking about vector sets instead of their Gram matrices while speaking of $E_p(A)$.

In Section 4, we prove that the $p$-frame energy for unit vectors in the plane is minimized by repeated orthonormal bases for any number of vectors if $p \in [1, 1.3]$. In Section 5 we discuss possible generalizations of the results of the paper and motivations behind them.

2. Auxiliary problems and tight frames

Bukh and Cox in [3] introduced a method for deriving new packing bounds for projective codes. In our related approach we use the notion of a *tight frame*. A tight frame is a finite collection of vectors $\{y_i\}_{i=1}^N$ in $\mathbb{R}^d$ with the property that,

$$\sum_{i=1}^N \langle x, y_i \rangle^2 = A \|x\|^2,$$

holds, for any $x \in \mathbb{R}^d$ and some $A > 0$.

Using the tight frame condition (1) and comparing coefficients for all $d$ components of $x$, one can conclude that $\sum_{i=1}^N \|y_i\|^2 =Cd$. It is convenient to use the normalization $A = \frac{1}{d}$ for frames as above, so that the Hilbert-Schmidt norm of the $d \times N$ matrix $Y$, with column vectors $\{y_i\}_{i=1}^N$, is normalized to be 1,

$$\|Y\|_{HS}^2 = \sum_{i=1}^N \|y_i\|^2 = 1.$$

In the next two lemmas we collect instruments for computing new lower bounds for discrete $p$-frame energies. The first makes a connection between kernels of matrices and associated tight frames. We introduce the notation $f_{c,p}(t) = \left(\frac{t}{c+t}\right)^\frac{p}{2}$, to be used in the second lemma. We also introduce the optimization problem,

$$M(c,p,N) = \min \left\{ \sum_{i=1}^N f_{c,p}(t_i) \left| \sum_{i=1}^N t_i = 1, \ t_i \in [0,c) \right. \right\},$$

where $p > 0$ and $c > 1/N$. Clearly, this optimization problem is properly defined.

**Lemma 2.1.** For any real $N \times N$ matrix $A$ of rank $d$, $N \geq d + 1$, with unit diagonal elements, there exists a tight frame $\{y_1, y_2, \ldots, y_N\} \subset \mathbb{R}^{N-d}$ with the frame constant $\frac{1}{\sqrt{1-d}}$ such that Ker $A$ consists of all vectors of the form $(\langle y, y_1 \rangle, \ldots, \langle y, y_N \rangle)$ with $y \in \mathbb{R}^{N-d}$.

**Proof.** Ker $A$ is $(N-d)$-dimensional so there is a linear mapping $L : \mathbb{R}^{N-d} \rightarrow \mathbb{R}^N$ whose image is Ker $A$. For each of $N$ components, $L$ is a linear functional so it may be represented as $L_i(y) = \langle y, z_i \rangle$. We note that for any non-singular mapping $D : \mathbb{R}^{N-d} \rightarrow \mathbb{R}^{N-d}$, the image of the mapping $L \circ D$ is Ker $A$ as well. The quadratic form $\sum_{i=1}^N \langle y, z_i \rangle^2$ is positive definite,
and so by choosing a suitable $D$, we can transform $\{z_1, z_2, \ldots, z_N\}$ into $\{y_1, y_2, \ldots, y_N\}$ so that $\sum_{i=1}^N \langle y_i, y_i \rangle^2 = \frac{1}{N-d} \langle y, y \rangle$. So, we obtain the condition 1, and $\{y_i\}_{i=1}^N$ is a tight frame. \qed

The construction in Lemma 2.1 is due to Bukh and Cox [3] who used it for obtaining new packing bounds for projective codes. It also can be interpreted as a tight frame representative of a Gale dual to the matrix $A$ (see [12] for more details about this interpretation).

Lemma 2.2. For any real $N \times N$ matrix $A$ of rank $d$, $N \geq d+1$, with unit diagonal elements,

$$E_p(A) \geq M \left( \frac{1}{N-d}, p, N \right)$$

if $1 \leq p \leq 2$,

$$E_p(A) \geq (N-1)^{1-\frac{p}{2}} M \left( \frac{1}{N-d}, p, N \right)$$

if $p \geq 2$.

Proof. By Lemma 2.1, there exists a tight frame $\{y_1, y_2, \ldots, y_N\} \subset \mathbb{R}^{N-d}$ such that $\text{Ker} A$ is the set of all vectors $(\langle y, y_1 \rangle, \ldots, \langle y, y_N \rangle)$ for some $y \in \mathbb{R}^{N-d}$ and $\sum_{i=1}^N |y_i|^2 = 1$. Taking $y = y_i$ and using the kernel condition for row $i$, we get

$$\langle y_i, y_i \rangle + \sum_{j \neq i} A_{ij} \langle y_i, y_j \rangle = 0.$$

Then for any $1 \leq i \leq N$,

$$\langle y_i, y_i \rangle \leq \sum_{j \neq i} |A_{ij}| |\langle y_i, y_j \rangle| \leq \left( \sum_{j \neq i} |A_{ij}|^p \right)^{\frac{1}{p}} \left( \sum_{j \neq i} |\langle y_i, y_j \rangle|^q \right)^{\frac{1}{q}},$$

by Hölder’s inequality for $\frac{1}{p} + \frac{1}{q} = 1$ (for $q = \infty$, $\left( \sum_{j \neq i} |\langle y_i, y_j \rangle|^q \right)^{\frac{1}{q}} = \max_{j \neq i} |\langle y_i, y_j \rangle|$).

By monotonicity of norms $\|\cdot\|_p$ in $p$ and Hölder’s inequality (for vectors in $\mathbb{R}^{N-1}$), $\|x\|_q \leq \|x\|_2$ for $q \geq 2$, while $\|x\|_q \leq (N-1)^{\frac{1}{2} - \frac{1}{2}} \|x\|_2$ when $q \leq 2$. Hence,

$$\langle y_i, y_i \rangle \leq \left( \sum_{j \neq i} |A_{ij}|^p \right)^{\frac{1}{p}} \left( \sum_{j \neq i} |\langle y_i, y_j \rangle|^2 \right)^{\frac{1}{2}} \text{ if } p \leq 2, \text{ and},$$

$$\langle y_i, y_i \rangle \leq (N-1)^{\frac{1}{2} - \frac{1}{p}} \left( \sum_{j \neq i} |A_{ij}|^p \right)^{\frac{1}{p}} \left( \sum_{j \neq i} |\langle y_i, y_j \rangle|^2 \right)^{\frac{1}{2}} \text{ if } p \geq 2.$$

At this point we use the tight frame condition for $y_i$, i.e. $\sum_{j \neq i} |\langle y_i, y_j \rangle|^2 = \frac{1}{N-d} \langle y_i, y_i \rangle - \langle y_i, y_i \rangle^2$, and denote $\langle y_i, y_i \rangle$ by $t_i$ to arrive finally at the inequalities:

$$\left( \sum_{j \neq i} |A_{ij}|^p \right)^{\frac{1}{p}} \geq \left( \frac{t_i}{N-d - t_i} \right)^{\frac{1}{2}} \text{ if } p \leq 2, \text{ and},$$

$$\left( \sum_{j \neq i} |A_{ij}|^p \right)^{\frac{1}{p}} \geq (N-1)^{\frac{1}{p} - \frac{1}{2}} \left( \frac{t_i}{N-d - t_i} \right)^{\frac{1}{2}} \text{ if } p \geq 2.$$
Taking powers, noting $\sum_{i=1}^{N} t_i = 1$, and summing these inequalities over all $i$, we obtain the conclusion of the lemma. \hfill \Box

3. New lower bounds for the $p$-frame energy

As a first application of Lemma 2.2, we give a new proof of the result from \cite{17} (also Proposition 3.1 in \cite{7}).

**Proposition 3.1.** For any $p \geq 2$ and any real $N \times N$ matrix $A$ of rank $d$, $N \geq 2$, with unit diagonal elements,

$$E_p(A) \geq N(N-1) \left( \frac{N-d}{d(N-1)} \right)^{\frac{p}{2}}. $$

**Proof.** For $N = d$, the right-hand side is 0. We assume $N \geq d + 1$ for the rest of the proof.

For $p \geq 2$, $f_{c,N}(t) = \left( \frac{t^{\frac{p}{2}}}{c^{\frac{p}{2}}} \right)$ is convex on $[0, c)$. Due to Jensen’s inequality, this implies

$$M \left( \frac{1}{N-d}, p, N \right) = N f_{\frac{1}{N-d},N} \left( \frac{1}{N} \right) = N \left( \frac{N-d}{d} \right)^{\frac{p}{2}}. $$

Together with Lemma 2.2 this completes the proof. \hfill \Box

It is straightforward to check that when $p > 2$, the only case in which the inequality of Proposition 3.1 is exact, holds when $Y = \{y_1, \ldots, y_N\}$ is a tight frame in $\mathbb{R}^{N-d}$ which satisfies that $|y_i|$ is constant for all $i$ and $|\langle y_i, y_j \rangle|$ is also constant for all $i \neq j$. In other words equality holds in Proposition 3.1 if and only if $Y$ is an equiangular tight frame (ETF). In view of $Y$ as an ETF, matrix $A$ is then the Gram matrix of the $d$-dimensional ETF known as the Naimark complement or Gale dual to $Y$ (see, for instance, \cite{4} and \cite{6} for more details about Naimark complements and Gale duality of equiangular tight frames).

There are several interesting properties of ETFs but the two most fundamental are that they are precisely the equality achieving systems of vectors for the Welch bound, and the maximum size $N$ of such systems is limited by Gerzon’s bound.

The Welch bound gives a lower bound for the coherence of a system of unit vectors namely, for unit vectors $\{\varphi_i\}_{i=1}^{N} \subset \mathbb{R}^d$ or $\mathbb{C}^d$,

$$\max_{i \neq j} |\langle \varphi_i, \varphi_j \rangle| \geq \sqrt{\frac{N-d}{d(N-1)}}. $$

This bound is one example of several other bounds which limit how spread out the one-dimensional subspaces corresponding to each vector may be \cite{22}; see \cite{15} for similar bounds and their derivation from a linear programming approach.

Gerzon’s bound \cite{14} limits the size of an ETF, and states in the real and complex case respectively,

$$N \leq \binom{d+1}{2} \quad \text{and} \quad N \leq d^2. $$

We call ETFs attaining this bound maximal. Maximal ETFs in real spaces $\mathbb{R}^d$ may exist only when $d = 1, 2, 3$ or $(2m+1)^2 - 2$ for some natural $m$. There are known maximal ETFs for $d = 1, 2, 3, 7, 23$ only. More details on the existence of maximal ETFs and the improvement of Gerzon’s bound for $d \neq (2m+1)^2 - 2$ are available in \cite{11}. The existence of complex maximal ETFs is the famous open conjecture \cite{25}. For maximal ETFs, the Welch bound must hold as
well so their coherence must satisfy $\alpha^2 = \frac{1}{d+2}$ in the real case and $\alpha^2 = \frac{1}{2}$ in the complex case. We discuss maximal ETFs further in connection with Theorem 5.1 in Section 5.

Using the lemmas from the previous section, we now give another observation on the relation between optimizing $E_p(A)$ and the problem $M(c, p, N)$.

**Theorem 3.1.** For any $1 \leq p < 2$ and any real $N \times N$ matrix $A$ of rank $d$ with unit diagonal elements,

$$E_p(A) \geq \frac{2(N - d)}{p^2(2 - p)^{\frac{d+2}{2}}}.$$  

**Proof.** For $N = d$, the right-hand side is 0. We assume $N \geq d + 1$ for the rest of the proof.

$f_{\frac{1}{N-d}, N}(t)$ is minimized on $(0, \frac{1}{N-d})$ at $t_0 = \frac{2-p}{2(N-d)}$. Hence,

$$f_{\frac{1}{N-d}, N}(t) \geq \frac{2(N - d)}{p^2(2 - p)^{\frac{d+2}{2}}} t, \quad \text{and so,}$$

$$M \left( \frac{1}{N-d}, p, N \right) \geq \frac{2(N - d)}{p^2(2 - p)^{\frac{d+2}{2}}} \sum_{i=1}^{N} t_i = \frac{2(N - d)}{p^2(2 - p)^{\frac{d+2}{2}}}.$$  

The final bound then follows from Lemma 2.2. \hfill \Box

When taking $p = 1$ in Theorem 3.1, we get $E_1(A) \geq 2(N - d)$. For $N$ in the range $d + 1 \leq N \leq 2d$, we thus obtain the bound conjectured in [16] from Theorem 3.1. We formulate it here as a simple statement about angles between lines in Euclidean spaces where it is understood that such angles are restricted to lie in $[0, \pi/2]$.

**Corollary 3.1.** The sum of cosines of all pairwise angles between $N$ lines in $\mathbb{R}^d$ is at least $N - d$. For $N \in [d, 2d]$, the bound is sharp and the unique minimizer is the set of $N$ lines forming a repeated orthonormal basis.

As hinted in the discussion [16], Corollary 3.1 may be proven by induction. For the sake of completeness, we provide such a proof here as well.

*Alternative proof of Corollary 3.1.* We choose a unit vector in the direction of each of the $N$ lines and construct an $N \times N$ Gram matrix for the chosen vectors. The matrix has rank no greater than $d$ so, by the Gershgorin circle theorem, any $(d + 1) \times (d + 1)$ diagonal minor of the matrix will have at least one row whose sum of absolute values of non-diagonal entries is at least 1. The inductive step consists then in finding one row like this and using the inductive hypothesis for the $(N - 1) \times (N - 1)$ diagonal minor obtained by deleting this row and the column symmetric to it from the matrix. \hfill \Box

We do not know how to extend this short proof to a more general problem of finding lower bounds for the 1-frame energy of matrices that is covered by Theorem 3.1. The proof of Corollary 3.1 does not seem to work for non-symmetric matrices. We also note that Theorem 3.1 implies the same bound $2(N - d)$ for $E_p(A)$ when $p \in (0, 1)$.

For the case of $N = d + 1$, Chen, Gonzales, Goodman, Kang, and Okoudjou [5] posed a conjecture for the minimum of the $p$-frame energy for all $p \in (0, 2)$. They conjectured that a global minimum is necessarily formed by $k + 1$ unit vectors whose endpoints form a regular $k$-dimensional simplex and $N - k - 1$ vectors that are pairwise orthogonal and orthogonal to the subspace of the regular simplex. In particular, their conjecture states that for $p < \frac{\ln 3}{\ln 2} \approx 1.58496$, the minimum is 2 and attained only on the repeated orthogonal basis with $d + 1$ vectors.
We study a more general problem of $d + m$ vectors, $1 \leq m < d$, and, using Lemma 2.2, prove that the repeated orthonormal basis minimizes $E_p$ for $p \in [1, p_m]$. In particular, for $m = 1$, our results confirm the conjecture from [5] for $p$ in the range $p \leq 2(\ln 3/\ln 2 - 1) \approx 1.16993$.

**Theorem 3.2.** For $p \in [1, 2 \log \frac{2m+1}{2m} / \log \frac{m+1}{m}]$, $1 \leq m < d$ and a real $(d + m) \times (d + m)$ matrix $A$ of rank $d$ with ones along the diagonal,

$$E_p(A) = \sum_{i \neq j} |A_{i,j}|^p \geq 2m.$$

The following lemma is used in the proof of Theorem 3.2.

**Lemma 3.1.** Set $\alpha = \frac{1}{m} \left( \frac{1}{2} - \frac{p}{4} \right)$. For $p \in [1, 2]$, $M(\frac{1}{m}, p, N)$ is minimized for $t_j$ of the form

(i) $t_1 = \ldots = t_k = \frac{1}{k}, \ t_{k+1} = \ldots = t_n = 0$, where $\frac{1}{k} \geq \alpha$, or,

(ii) $t_1 = \ldots = t_k = x, \ t_{k+1} = 1 - kx, \ t_{k+2} = \ldots = t_N = 0$, where $x \geq \alpha$, $0 < 1 - kx < \alpha$.

**Proof.** Computing the second derivative of $f_{1/m,p}(t)$, we see it is concave on $[0, \alpha]$ and convex on $[\alpha, 1)$ where $\alpha = \frac{1}{m} \left( \frac{1}{2} - \frac{p}{4} \right)$. All $t_i$ lying in $[0, \alpha]$ may be moved to the endpoints of the interval, except for at most one number, while keeping their sum constant and minimizing the sum of values of $f_{1,p}$ (this follows, for instance, from the Karamata inequality, see [13, pg 89]). After this we may apply Jensen’s inequality for all numbers from $[\alpha, 1)$ and assume they are all equal. The resulting minimizer is then necessarily one of the two types: 1) $t_1 = \ldots = t_k = \frac{1}{k}, \ t_{k+1} = \ldots = t_N = 0$, where $\frac{1}{k} \geq \alpha$, 2) $t_1 = \ldots = t_k = x, \ t_{k+1} = 1 - kx, \ t_{k+2} = \ldots = t_N = 0$, where $x \geq \alpha$ and $0 < 1 - kx < \alpha$. \hfill $\square$

We now follow with a proof of Theorem 3.2.

**Proof.** Set $p_m = 2 \log \frac{2m+1}{2m} / \log \frac{m+1}{m}$ and $q_m = \frac{p_m}{m}$. Clearly, it is sufficient to prove the lower bound for $p = p_m$ only. We use Lemma 2.2 and show that $M(\frac{1}{m}, p, N) \geq 2m$. Consider the first case in the above lemma, so that

$$t_1 = \ldots = t_k = \frac{1}{k}, \ t_{k+1} = \ldots = t_n = 0,$$

where $\frac{1}{k} \geq \alpha = \frac{1}{m} \left( \frac{1}{2} - \frac{p_m}{4} \right)$. In this case, we need to minimize the value

$$kf_{1/m,p_m}\left( \frac{1}{k} \right) = k \left( \frac{m}{k - m} \right) \frac{p_m}{2}.$$

The real function

$$F_m(x) = x \left( \frac{m}{x - m} \right) \frac{p_m}{2}$$

for $x > m$ has exactly one local minimum. The degree $p_m$ was specifically chosen so that $F_m(2m) = F_m(2m + 1) = 2m$. Then by Rolle’s theorem, the local minimum of $F_m(x)$ lies in $[2m, 2m + 1]$. The minimum of $F_m(x)$ for natural values of $x$, $x > m$, is, therefore, attained on $2m$ and $2m + 1$ and is equal to $2m$.

In the second case, $x < \frac{1}{k}$ and $x \geq \alpha = \frac{1}{m} \left( \frac{1}{2} - \frac{p_m}{4} \right)$. It is straightforward to show that $p_m < \frac{4m+2}{4m+1}$ for all natural $m$. Subsequently $k < \frac{1}{\alpha} < 4m + 1$ so that $k$ can take (integer) values only in $[m, 4m]$. To show $E_{p_m} \geq 2m$ it suffices then to show for all $m \leq j \leq 4m$, and all $x$ in $I = (\frac{1}{j+1}, \frac{1}{j})$ that the function

$$g_j(x) = j \left( \frac{mx}{1 - mx} \right)^{q_m} + \left( \frac{m(1 - jx)}{1 - m(1 - jx)} \right)^{q_m},$$
satisfies $g_j(x) \geq 2m$. This will be demonstrated using properties specific to $g_j(x)$, namely that the function has at most one critical point, $g'_j(x) = 0$, inside the interval $I$. Taking derivatives,

$$g'_j(x) = q_m jm \left[ \left( \frac{mx}{1 - mx} \right)^{q_m - 1} \frac{1}{(1 - mx)^2} - \left( \frac{m(1 - jx)}{1 - m(1 - jx)} \right)^{q_m - 1} \frac{1}{(1 + m(-1 + jx))^2} \right],$$

so that $g'_j(x) = 0$ gives

$$\left( \frac{x(1 + m(-1 + jx))}{(1 - mx)(1 - jx)} \right)^{q_m - 1} \frac{1}{(1 - mx)(1 - jx)} = \frac{(1 - mx)^2}{(1 + m(-1 + jx))^2}$$

and

$$\left( \frac{x(1 + m(-1 + jx))}{(1 - mx)(1 - jx)} \right)^{q_m + 1} \frac{1 + m(-1 + jx)}{1 - mx} = \frac{x^2}{(1 - jx)^2} \frac{1}{q_m + 1} \frac{2}{1 - \frac{2}{q_m + 1}}.$$

Calling the function on the left in the above expression $f(x)$ and the function on the right $g(x)$,

$$f''(x) = \frac{2j(1 + j - m)m^2}{(1 + m(-1 + jx))^3} > 0 \text{ on } I,$$

while letting $\gamma = 1 - \frac{2}{q_m + 1}$,

$$g''(x) = \frac{\gamma(\frac{x}{1 - jx})^\gamma(\gamma - 1 + 2jx)}{x^2(1 - jx)^2} < 0 \text{ on } I,$$

since $\gamma < 0$. Thus $f(x)$ is convex on $I$, while $g(x)$ is concave on $I$. Since $f(\frac{1}{j+1}) = g(\frac{1}{j+1})$ and $f'(\frac{1}{j+1}) \leq g'(\frac{1}{j+1})$, when $j < 4m$, it must be the case then that $f(x) = g(x)$ for exactly one point $x \in I$, $(x \neq \frac{1}{j+1}, \frac{1}{j})$. Note that when $j = 4m$ there are no critical points in $I$. Now,

$$g'_j \left( \frac{1}{j + 1} \right) = 0 \text{ and } \lim_{x \to \frac{1}{j}} g'_j(x) = -\infty.$$

Thus the critical points correspond to local maxima of $g_j(x)$ and it suffices to check the value of $g_j(x)$ at the endpoints in $I$ for each $m \leq j \leq 4m$ to establish the desired lower bound. These values are

$$g_j \left( \frac{1}{j + 1} \right) = (1 + j) \left( \frac{m}{1 + j - m} \right)^{q_m} = F_m(j + 1), \text{ and } g_j \left( \frac{1}{j} \right) = j \left( \frac{m}{j - m} \right)^{q_m} = F_m(j),$$

where $F_m$ is the function defined in equation (4). The minimal value of $F_m(x)$ on natural numbers, $x > m$, as we established earlier, is precisely $2m$. □

Following the proofs of Theorem 3.2 and Lemma 2.2 it is easy to check that the only minimizer is the repeated orthonormal basis. Theorem 3.2 was first announced by the second author in [18].
4. Minimizing energy in two dimensions

In this section, we study the problem of minimizing the $p$-frame energy for collections of unit vectors in the plane. In particular, we show that the repeated orthonormal basis is the minimizer for $p \leq 1.3$. As one of the instruments for our proof, we use the solution of the Fejes Tóth problem mentioned in Section 1.

**Theorem 4.1.** Let $x_1, x_2, \ldots, x_N$ be unit vectors in the plane. Then,

$$\sum_{i,j=1}^{N} \arccos |\langle x_i, x_j \rangle| \leq \frac{\pi N^2}{4}, \text{ if } N \text{ is even},$$

$$\sum_{i,j=1}^{N} \arccos |\langle x_i, x_j \rangle| \leq \frac{\pi(N^2 - 1)}{4}, \text{ if } N \text{ is odd}.$$

Theorem 4.1 was proven in [9]. Several alternative proofs were also obtained in [1].

**Theorem 4.2.** Let $A$ be a Gram matrix of $N$ unit vectors in the plane. Then for $p \in (0,1.3]$, $E_p(A) \geq N(N - 2)/2$ if $N$ is even and $E_p(A) \geq (N - 1)^2/2$ if $N$ is odd.

**Proof.** Assume $A$ is the Gram matrix of unit vectors $x_1, x_2, \ldots, x_N$ in the plane.

For any $p \in [0,2]$, the lower bound on $E_p$ for even $N$ follows immediately from the fact that a repeated orthonormal basis is one of the minimizers of $E_2$ when the number of vectors is divisible by the dimension (see [7, 20, 21]).

It is sufficient to prove the lower bound for $p = 1.3$ so we consider this case only for the rest of the proof. For odd $N$, we split our proof into two parts: 1) angles between each pair of vectors are sufficiently far from $\pi/2$; 2) there are vectors that are almost orthogonal. For the first case, we assume that all angles $\arccos |\langle x_i, x_j \rangle|$, $1 \leq i,j \leq N$, are no greater than 1.34. It is straightforward to check that for any $t \in [0,1.34]$,

$$\cos^{1.3} t \geq \frac{2}{\pi} \left( \frac{\pi}{2} - t \right).$$

Summing these inequalities for all $t = \arccos |\langle x_i, x_j \rangle|$, $1 \leq i,j \leq N$, $i \neq j$, we obtain

$$E_{1.3}(A) \geq N^2 - N - \frac{2}{\pi} \sum_{i,j=1}^{N} \arccos |\langle x_i, x_j \rangle|.$$

Using the solution of the Fejes Tóth problem from Theorem 4.1 we conclude

$$E_{1.3}(A) \geq N^2 - N - \frac{2}{\pi} (N^2 - 1)\frac{\pi}{4} = \frac{(N - 1)^2}{2}.$$

For the second case, we assume that the largest angle among $\arccos |\langle x_i, x_j \rangle|$, $1 \leq i,j \leq N$, is at least 1.34. Without loss of generality, let one of such angles be $\arccos |\langle x_1, x_2 \rangle|$. Our proof will be by induction on odd numbers $N$. The statement of the theorem is clearly true for $N = 1$. Let $N$ be an odd number greater than 1. Now we will show that for any $i, 3 \leq i \leq N$, $|\langle x_1, x_i \rangle|^{1.3} + |\langle x_2, x_i \rangle|^{1.3} \geq 1$.

We can always switch a vector to its opposite without changing the total energy so we may assume that $x_i$ lies in the angle formed by $x_1$ and $x_2$. We assume the angle between $x_1$ and $x_2$ is $\theta$ and $x_i$ forms the angles of $\phi$ and $\theta - \phi$ with $x_1$ and $x_2$, respectively. Note that both $\phi$ and $\theta - \phi$ must be less than $\pi/2$, otherwise one of them is closer to $\pi/2$ than $\theta$. Without loss of generality, $\phi \leq \theta/2$. There are two possible options: $\theta \leq \pi/2$ and $\theta > \pi/2$. 


For the first option,
\[ |\langle x_1, x_i \rangle|^{1.3} + |\langle x_2, x_i \rangle|^{1.3} = \cos^{1.3} \phi + \cos^{1.3}(\vartheta - \phi) \geq \cos^{1.3} \phi + \cos^{1.3} \left( \frac{\pi}{2} - \phi \right) \geq 1. \]

For the second option, \( \vartheta \leq \pi - 1.34 \) because \( \vartheta = \pi - \arccos |\langle x_1, x_2 \rangle| \). The angle \( \vartheta \) is the one closest to \( \pi/2 \) among all angles formed by the vectors. In particular, \( \vartheta - \phi \) cannot be closer to \( \pi/2 \) so \( \vartheta - \phi \leq \pi - \vartheta \). This condition can be rewritten as \( \frac{\pi - \vartheta}{2} \geq \vartheta - \phi \). For the next step we try to minimize \( \cos^{1.3} \phi + \cos^{1.3}(\vartheta - \phi) \) by keeping \( \phi \) intact and increasing \( \vartheta - \phi \) as much as possible while preserving the conditions \( \vartheta - \phi \leq \pi - \phi - 1.34 \) and \( \vartheta - \phi \leq \frac{\pi - \vartheta}{2} \). While increasing \( \vartheta \), at some moment we reach the point when one of these two inequalities becomes a precise equality. These two possibilities can be described by the two cases depending on the value of \( \phi \).

For the first case, assume \( \frac{\pi - \vartheta}{2} > 1.34 \), i.e. \( \phi < \pi - 2.68 \). Then
\[ \cos^{1.3} \phi + \cos^{1.3}(\vartheta - \phi) \geq \cos^{1.3} \phi + \cos^{1.3} \left( \frac{\pi - \phi}{2} \right) = \cos^{1.3} \phi + \sin^{1.3} \frac{\phi}{2}. \]
The function \( \cos^{1.3} \phi + \sin^{1.3} \frac{\phi}{2} \) is at least 1 for \( \phi \in [0, \pi - 2.68] \) so the first case is covered.

For the second case, we assume \( \frac{\pi - \vartheta}{2} \leq 1.34 \), i.e. \( \phi \geq \pi - 2.68 \). We know that \( \phi \geq \vartheta/2 \leq \frac{\pi - 1.34}{2} \). Using the inequality \( \vartheta < \pi - 1.34 \) again we get that \( \vartheta - \phi < \pi - 1.34 - \phi \). This implies
\[ \cos^{1.3} \phi + \cos^{1.3}(\vartheta - \phi) \geq \cos^{1.3} \phi + \cos^{1.3}(\pi - 1.34 - \phi). \]
The function \( \cos^{1.3} \phi + \cos^{1.3}(\pi - 1.34 - \phi) \) is at least 1 for \( \phi \in [\pi - 2.68, \frac{\pi - 1.34}{2}] \). Overall, we conclude that \( |\langle x_1, x_i \rangle|^{1.3} + |\langle x_2, x_i \rangle|^{1.3} = \cos^{1.3} \phi + \cos^{1.3}(\vartheta - \phi) \geq 1. \)

Then, by the induction hypothesis for unit vectors \( x_3, \ldots, x_N \),
\[ E_{1.3}(A) \geq \frac{(N - 3)^2}{2} + 2 \sum_{i=3}^{N} (|\langle x_2, x_i \rangle|^{1.3} + |\langle x_1, x_i \rangle|^{1.3}) \geq \frac{(N - 3)^2}{2} + 2(N - 2) = \frac{(N - 1)^2}{2}. \]

We do not know how to extend this proof to the general case of matrices of rank 2. The value \( p = 1.3 \) is not the best possible. One can impose all conditions necessary for the proof of Theorem 4.2 to work and optimize for \( p \). The numerical value obtained this way is approximately 1.317.

5. Discussion

Recently, the conjecture mentioned above from the paper [5] was proved true in [24]. In particular, the range that the orthonormal basis plus a vector minimizes in Theorem 3.2 can be increased to the value \( p = \log 3/ \log 2 \). The behavior of the other maximal values of \( p \) that similar constructions are expected to minimize for is suggested in the following conjecture (appearing first in [18]).

Conjecture 5.1. Let \( N = m + kd \) points be given in \( S^{d-1} \), with \( 1 \leq m < d \), \( d \geq 2 \), and a Gram matrix \( A \in \mathbb{R}^{N \times N} \). Then there is a value of \( p_0 \), independent of dimension \( d \) and excess \( m \), such that the repeated orthonormal sequence \( \{e_{j \mod d}\}_{j=1}^N \) minimizes \( E_p \) over all size \( N \) systems of unit vectors (with value \( E_p(A) = d(k^2 - k) + 2k \)) for \( p < p_0 \) and the minimum value of \( E_p(A) \) satisfies \( E_p(A) < d(k^2 - k) + 2k \) when \( p > p_0 \). Further \( p_0 = p_0(k) \) satisfies \( p_0(k) \to 2 \) as \( k \to \infty \).
Theorem 4.2, can be interpreted as an improvement to Theorem 3.2 from the previous section, and partial progress towards the above conjecture. The parameter $p_0 = 1.3$ in Theorem 4.2 is the same for all values of $k$. It might be interesting to find an elementary argument which showed that $p_0(k) \to 2$ in dimension 2, let alone generally.

Theorem 3.2 and Conjecture 5.1 appear to be examples of a more general phenomenon. The direct analogs/extensions of orthonormal bases are regular simplices in the non-projective setting and maximal ETFs in projective spaces. In the first case, the analogous potential function is $f^{\Delta}_{d,p}(t) = |t + \frac{1}{d}|^p$. In the second case, one can view the potential function $f(t) = |t|^p$ as an instance of the more general function $f_{\alpha,p}(t) = |t^2 - \alpha^2|^p$, and the orthonormal basis as an example of an equiangular tight frame ($|\langle x, y \rangle| = \alpha$, for $x \neq y$ with coherence $\alpha = 0$).

The problem of minimizing the energies $E^{\Delta}_{d,p}$ and $E_{\alpha,p}$, associated with $f^{\Delta}_{d,p}$ and $f_{\alpha,p}$, respectively, for $p$ close to 0 might be expected to pick out repeated regular simplices and repeated ETFs with coherence $\alpha$. We conjecture generally:

**Conjecture 5.2.**

(i) Let $A \in \mathbb{R}^{N \times N}$ be the Gram matrix of $N = m + k(d + 1)$ points given in $\mathbb{S}^{d-1}$, with $1 \leq m < d + 1$, $d \geq 2$, and let $\{\varphi_i\}_{i=1}^m \subset \mathbb{R}^d$ be the maximal regular simplex in $\mathbb{S}^{d-1}$. Then there is a value of $p_0 > 0$, such that the repeated regular simplex $\{\varphi_{\text{mod } t}\}_{i=1}^N$ minimizes $E^{\Delta}_{d,p}$ over all size $N$ systems of unit vectors for $p < p_0$.

(ii) Let $A \in \mathbb{R}^{N \times N}$ be the Gram matrix of $N = m + kl$ points given in $\mathbb{S}^{d-1}$, with $1 \leq m < l$, $d \geq 2$, and $l$ the size of an ETF with coherence $\alpha$, $\{\varphi_{\text{mod } t}\}_{i=1}^N \subset \mathbb{R}^d$. Then there is a value of $p_0 > 0$, such that the repeated ETF sequence $\{\varphi_{\text{mod } t}\}_{i=1}^N$ minimizes $E_{\alpha,p}$ over all size $N$ systems of unit vectors for $p < p_0$.

Conjectures 5.1 and 5.2 emphasize the possibility of repeated configurations minimizing $p$-frame or $p$-frame-type energies among sets of $N$ points for all large enough $N$. This property may be seen as a strong version of stability of an optimizing set.

The collections of unit vectors $\Phi = \{\varphi_{\text{mod } t}\}_{i=1}^N \subset \mathbb{R}^d$ that, for all $N \geq l$, have a repeated set $\{\varphi_{\text{mod } t}\}_{i=1}^N$ minimizing the energy defined by the potential function $f$ among all $N \times N$ Gram matrices $A$, also satisfy that the uniform distribution over $\Phi$ must solve the problem,

$$\min_{\mu \in \mathcal{P}(\mathbb{S}^{d-1})} I_f(\mu) = \min_{\mu \in \mathcal{P}(\mathbb{S}^{d-1})} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{S}^{d-1}} f(\langle x, y \rangle) d\mu(x) d\mu(y),$$

where $\mathcal{P}(\mathbb{S}^{d-1})$ are Borel probability measures, $\mu(\mathbb{S}^{d-1}) = 1$. For the case of $p$-frame energies, tight designs are examples of configurations such that uniform distributions over them minimize $I_f(\mu)$ over ranges of $p$ [2]. This behavior is slightly different from that conjectured above, as repeated tight designs of size $l$ can only be expected to minimize the discrete energies when $N = kl$, generally.

The existence of a $p_0$ in these conjectures might be expected in connection with ideas from the field of compressed sensing. One should expect that, as $p \to 0$, the solution to this minimization problem is a repeated ETF (as upon vectorizing the Gram matrix, and considering the difference, the sparsest difference arises this way), but Conjecture 5.2 strengthens this to say that the solution for $p$ sufficiently small is also a repeated ETF.

In connection with these observations we collect further support for the above conjectures for maximal ETFs and the regular simplex, by showing they minimize the associated continuous energies (5). The second part of the below theorem holds also with the coherence replaced with the corresponding value on a complex maximal ETF, these being known alternatively as symmetric informationally complete positive operator-valued measures (SIC-POVMs, see [10] for more details). We give the proof only in the real case, but the same type of argument applies in the complex case.
Theorem 5.1. The following statements hold:

(i) The uniform distribution over the vertices of a regular simplex \( \{ \varphi_i \}_{i=1}^{d+1} \subset S^{d-1} \) minimizes the continuous energy \( I_{d,p}^\Delta \) for \( f_{d,p}(t) = |t + \frac{1}{d}|^p \) for all \( p \in (0,2] \).

(ii) Whenever a maximal ETF exists, \( \{ \varphi_i \}_{i=1}^{M} \subset S^{d-1} \) (with coherence \( \alpha^2 = \frac{1}{d+2} \)), the uniform distribution over its points minimizes the continuous energy \( I_{\alpha,p} \) for \( f_{\alpha,p}(t) = |t^2 - \alpha^2|^p \) for all \( p \in (0,2] \).

Proof. Note that the inequalities

\[
|t + \frac{1}{d}|^p \geq \left| \frac{t + 1}{1 + \frac{1}{d}} \right|^2 \quad \text{and} \quad \left| \frac{t^2 - \frac{1}{d+2}}{1 - \frac{1}{d+2}} \right|^p \geq \left| \frac{t^2 - 1}{1 - \frac{1}{d+2}} \right|^2,
\]

hold for \( t \in [-1,1] \). The first inequality implies \( I_{d,p}^\Delta \geq (1 + \frac{1}{d})^p - 2^p I_{d,2}^\Delta \) and the equality holds for the uniform distribution over the regular simplex. The second inequality implies \( I_{\alpha,p} \geq (1 - \frac{1}{d+2})^p - 2^p I_{\alpha,2} \) and the equality holds for the uniform distribution over a maximal ETF. It is sufficient then to prove the theorem for \( p = 2 \).

For the proof in the case \( p = 2 \), we use the notion of positive definite functions on unit spheres. A function of inner products is positive definite on \( S^{d-1} \) if for any collection of points \( \{ x_i \}_{i=1}^{N} \subset S^{d-1} \) and sequence of complex numbers \( \{ c_i \}_{i=1}^{N} \subset \mathbb{C} \) it is true that

\[
\sum_{i,j=1}^{N} c_i \overline{c_j} f(\langle x_i, x_j \rangle) \geq 0.
\]

This condition implies that for any positive definite \( f \) and any measure \( \mu \) on \( S^{d-1} \),

\[
I_f(\mu) = \int_{S^{d-1}} \int_{S^{d-1}} f(\langle x, y \rangle) d\mu(x) d\mu(y) \geq 0.
\]

Positive definite functions on spheres were characterized by Schoenberg [19]. In particular, \( t, t^2 - \frac{1}{d}, \) and \( t^4 - \frac{6}{d+4} t^2 + \frac{3}{(d+2)(d+4)} \) (Gegenbauer polynomials of degrees 1, 2, 4) are positive definite functions on \( S^{d-1} \).

Since \( (t + \frac{1}{d})^2 = (t^2 - \frac{1}{d}) + \frac{2}{d} t + \frac{d+1}{d^2} \),

\[
I_{d,2}^\Delta(\mu) \geq \frac{d+1}{d^2}.
\]

It is easy to check that for the uniform distribution over the regular simplex, \( I_{d,2}^\Delta \) is precisely \( \frac{d+1}{d^2} \).

Using

\[
\left( t^2 - \frac{1}{d+2} \right)^2 = \left( t^4 - \frac{6}{d+4} t^2 + \frac{3}{(d+2)(d+4)} \right) + \frac{4(d+1)}{(d+2)(d+4)} \left( t^2 - \frac{1}{d} \right) + \frac{2(d+1)}{d(d+2)^2},
\]

we conclude that

\[
I_{\alpha,2}(\mu) \geq \frac{2(d+1)}{d(d+2)^2}.
\]

Again it is easy to check that the equality is attained on the uniform distribution over the maximal ETF. \( \Box \)
Using the design conditions, one can also show that the configurations from Theorem 5.1 are unique minimizers (up to the uniqueness of maximal ETFs in the second case) for the corresponding energies when $p \in (0, 2)$, similarly to how it is done in [2] for the general case of tight designs and $p$-frame energies.

6. Acknowledgments

The authors would like to thank Wei-Hsuan Yu for fruitful discussions, Boris Bukh and Chris Cox for sharing the manuscript and discussing their results, Kasso Okoudjou for talking about his and his coauthors’ research in this direction and informing about their conjectures. This paper is based upon work supported by the National Science Foundation under Grant No. DMS-1439786 while the first author was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Spring 2018 semester. The second author was supported in part by grant from the US National Science Foundation, DMS-1600693.

References

[1] D. Bilyk and R. Matzke. On the Fájes Tóth problem about the sum of angles between lines. Proc. Amer. Math. Soc., 147 (2019), 51–59. arXiv:1801.07837
[2] D. Bilyk, A. Glazyrin, R. Matzke, J. Park, and O. Vlasiuk. Optimal measures for $p$-frame energies on spheres. preprint. arXiv:1908.00885
[3] B. Bukh and C. Cox. Nearly orthogonal vectors and small antipodal spherical codes. preprint. arXiv:1803.02949
[4] P. G. Casazza, M. Fickus, D. G. Mixon, J. Peterson, and I. Smalyanau. Every Hilbert space frame has a Naimark complement. Journal of Mathematical Analysis and Applications, 406 (2013), 111–119. arXiv:1104.0810 MR 3062405
[5] X. Chen, V. Gonzales, E. Goodman, S. Kang and K. A. Okoudjou. Universal optimal configurations for the $p$-frame potentials. preprint. arXiv:1902.03505
[6] H. Cohn, A. Kumar, and G. Minton. Optimal simplices and codes in projective spaces. Geometry & Topology, 20 (2016), 1289–1357. arXiv:1308.3188 MR 3523059
[7] M. Ehler and K. A. Okoudjou. Minimization of the probabilistic $p$-frame potential. J. Stat. Plan. Inference, 142 (2012), 645–659. arXiv:1101.0140 MR 2853573
[8] L. Fejes Tóth. Über eine Punktverteilung auf der Kuge. Acta Math. Acad. Sci. Hungar., 10 (1959), 13–19. MR 0105654
[9] F. Fodor, V. Vígh, and T. Zarnócz. On the angle sum of lines. Arch. Math, 106 (2016), 91–100. MR 3451371
[10] C. A. Fuchs, M. C. Hoang, B. C. Stacey The SIC Question: History and State of Play. preprint. arXiv:1703.07901
[11] A. Glazyrin and W.-H. Yu. Upper bounds for $s$-distance sets and equiangular lines. Adv. Math 330 (2018), 810–833. arXiv:1611.09479 MR 3787558
[12] A. Glazyrin. Moments of isotropic measures and optimal projective codes. preprint. arXiv:1904.11159
[13] G. H. Hardy, J. E. Littlewood and, G. Pólya. Inequalities. Cambridge Mathematical Library, Cambridge University Press, Cambridge, (1988). MR 94909
[14] P. W. H. Lemmens and J. J. Seidel. Equiangular lines. Journal of Algebra 24 (1973), 494–512. MR 0307969
[15] V. Levenstein. Bounds of the maximal capacity of a code with a limited scalar product modulus. Dokl. Akad. Nauk SSSR 263 (1982), 1303–1308. MR 0653223
[16] The minimum of a sum of absolute values of inner products in $\mathbb{R}^d$. Electronic resource. https://mathoverflow.net/q/272692.
[17] O. Oktay. Frame quantization theory and equiangular tight frames. Ph. D. thesis, University of Maryland (2007).
[18] J. Park. Frame potentials and orthogonal vectors. To appear in: 2019 International Conference on Sampling Theory and Applications Proceedings (SampTA), IEEE. arXiv:1902.09443
[19] I. J. Schoenberg. Positive definite functions on spheres. Duke Math. J., 9 (1941), 96–108. MR 0005922
[20] V. M. Sidel’nikov. New estimates for the closest packing of spheres in $n$-dimensional Euclidean space. Mat. Sb., 24 (1974), 148–158. MR 0362060
[21] B. Venkov. Réseaux et designs sphériques. Réseaux euclidiens, designs sphériques et formes modulaires, L’Enseignement mathématique, 37 (2001), 10–86. MR 1878746
[22] L. Welch. Lower bounds on the maximum cross correlation of signals. IEEE Trans. Inform. Theory, 20 (1974), 397–399.
[23] J. Wen, D. Li, and F. Zhu. Stable recovery of sparse signals via $\ell_p$-minimization. Applied and Computational Harmonic Analysis, 38 (2015), 161–176. arXiv:1406.4328 MR 3273293
[24] Z. Xu and Z. Xu. The minimizers of the $p$-frame potential. preprint. arXiv:1907.10861
[25] G. Zauner. Grundzüge einer nichtkommutativen Designtheorie. PhD thesis, University of Vienna, 1999. Published in English translation: Int. J. Quantum Inf. 9 (2011), 445–507. MR 2931102

School of Mathematical and Statistical Sciences, The University of Texas Rio Grande Valley, Brownsville, TX 78520
E-mail address: alexey.glazyrin@utrgv.edu

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160
E-mail address: j.park@gatech.edu