Auto-tuning unit norm frames

Peter G. Casazza[a], Matthew Fickus[b], Dustin G. Mixon[c]

[a]*Department of Mathematics, University of Missouri, Columbia, Missouri 65211, USA*
[b]*Department of Mathematics and Statistics, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio 45433, USA*
[c]*Program in Applied and Computational Mathematics , Princeton University, Princeton, New Jersey 08544, USA*

Abstract

Finite unit norm tight frames provide Parseval-like decompositions of vectors in terms of redundant components of equal weight. They are known to be exceptionally robust against additive noise and erasures, and as such, have great potential as encoding schemes. Unfortunately, up to this point, these frames have proven notoriously difficult to construct. Indeed, though the set of all unit norm tight frames, modulo rotations, is known to contain manifolds of nontrivial dimension, we have but a small finite number of known constructions of such frames. In this paper, we present a new iterative algorithm—gradient descent of the frame potential—for increasing the degree of tightness of any finite unit norm frame. The algorithm itself is trivial to implement, and it preserves certain group structures present in the initial frame. In the special case where the number of frame elements is relatively prime to the dimension of the underlying space, we show that this algorithm converges to a unit norm tight frame at a linear rate, provided the initial unit norm frame is already sufficiently close to being tight. By slightly modifying this approach, we get a similar, but weaker, result in the non-relatively-prime case, providing an explicit answer to the Paulsen problem: "How close is a frame which is almost tight and almost unit norm to some unit norm tight frame?"

Keywords: frames, finite, tight, unit norm, frame potential, gradient descent

1. Introduction

Frames provide numerically stable methods for finding overcomplete decompositions of vectors, and are ubiquitous in signal processing applications [16, 17]. As explained below, *tight frames* and *unit norm* frames are particularly useful. However, it is difficult to construct frames which possess both of these properties simultaneously, called *unit norm tight frames* (UNTFs). In this paper, we present a new method for overcoming this difficulty, namely an iterative procedure which, when applied to a given finite unit norm frame, asymptotically produces a UNTF. To be precise, under the additional assumptions that the number of frame vectors is relatively prime to the dimension of the underlying space and that our initial unit norm frame is sufficiently close to being tight, we are able to show that our method, namely a gradient descent of the *frame potential*, converges to a UNTF at a linear rate. That is, from a tightness perspective, our algorithm takes a good unit norm frame and makes it perfect. As such, it can be viewed as a frame-theoretic analog of *Auto-Tune*$^{\text{TM}}$, the software commonly used in the music industry to perfect the pitch of lesser vocalists. Moreover, in the non-relatively-prime case, we can slightly modify our argument to yield an explicit answer to the *Paulsen problem* [2]:

"How close is a frame which is almost tight and almost unit norm to some UNTF?"

To make these notions precise, consider the *synthesis operator* of a sequence of vectors $F = \{f_n\}_{n=1}^N$ in a real or complex M-dimensional Hilbert space $\mathbb{H}_M$, namely $F : \mathbb{C}^N \to \mathbb{H}_M$, $Fg := \sum_{n=1}^N g(n) f_n$. That is, viewing $\mathbb{H}_M$ as $\mathbb{R}^M$ or $\mathbb{C}^M$, F is the $M \times N$ matrix whose columns are the f_n's. Note that here and throughout, we make no notational distinction between the vectors themselves and the synthesis operator they induce. The vectors F are said to be a

Email address: Matthew.Fickus@afit.edu (Matthew Fickus)

frame for $\mathbb{H}_M$ if there exists *frame bounds* $0 < A \leq B < \infty$ such that $A\|f\|^2 \leq \|F^*f\|^2 \leq B\|f\|^2$ for all $f \in \mathbb{H}_M$. In this finite-dimensional setting, having F be a frame is equivalent to having the f_n's span $\mathbb{H}_M$, necessitating $M \leq N$, with the optimal frame bounds A and B corresponding to the least and greatest eigenvalues of FF^*. In particular, F is a *tight frame* when $A = B$, that is, when $FF^* = A\mathrm{I}$. Tight frames are useful in applications, as they provide Parseval-like decompositions

$$f = \tfrac{1}{A}FF^*f = \tfrac{1}{A}\sum_{n=1}^{N}\langle f, f_n\rangle f_n, \quad \forall f \in \mathbb{H}_M, \tag{1}$$

despite the fact that the f_n's are not required to be independent. Indeed, the tightness condition $FF^* = A\mathrm{I}$ does not require the columns of F, that is, the f_n's, to be orthogonal, but rather, it requires the rows of F to be orthogonal and have equal norm $\sqrt{A}$. Meanwhile, F is a *unit norm* frame when $\|f_n\| = 1$ for all $n = 1, \ldots, N$. When a frame is both unit norm and tight—a UNTF—it breaks vectors into possibly redundant components of equal weight (1), with the tight frame constant A being the redundancy $\frac{N}{M}$. UNTFs are known to be exceptionally robust against additive noise and erasures [7, 12, 13, 14]. Unfortunately, UNTFs are also notoriously difficult to construct: we want $M \times N$ matrices F that have unit norm columns and orthogonal rows of equal squared-norm $\frac{N}{M}$. To be clear, UNTFs are known to exist for any $M \leq N$: one may either invoke the classical theory of *majorization* for matrices, or more simply, consider the *harmonic frame* obtained by truncating an $N \times N$ discrete Fourier transform (DFT) matrix [12]. Another technique is to build an operator with a flat spectrum using weighted DFT blocks; this *spectral tetris* method yields extremely sparse UNTFs [6]. However, these techniques only produce certain examples of UNTFs, while the set of all UNTFs, modulo rotations, contains nontrivial manifolds whenever $N > M + 1$ [10]. That is, these methods produce but a few samples from the continuum.

In this paper, we provide a new method for starting with a given frame and producing a nearby UNTF from it. Such techniques are very useful in real-world problems, as they allow one to take a given transform, carefully crafted to have certain application-specific properties without being tight and/or unit norm, and to correct, or *tune*, its algebraic properties while changing the transform itself as little as possible. In terms of mathematics, these techniques are important because they help in solving the Paulsen problem. To be precise, a compactness argument of D. Hadwin [2] shows that indeed, if a frame is sufficiently close to being both tight and unit norm, then it is, in fact, close to a UNTF. Current work on this problem therefore focuses on *how* close these UNTFs are, as well as developing practical schemes to obtain them. Unfortunately, finitely-iterative techniques using Givens rotations [8, 14] have, to this point, produced UNTFs that are not necessarily close to the originals.

More recent approaches to solving the Paulsen problem, namely that of [2] and the present method, rely upon the fact that given any frame F, it is straightforward to produce a unit norm frame from it: simply replace each f_n with $\frac{f_n}{\|f_n\|}$. Moreover, one can also convert any frame into a tight frame, provided one has the computational power to take the inverse square root of the frame operator: consider $(FF^*)^{-\frac{1}{2}}F$. However, combining these two operations—dividing by the root of the frame operator and then normalizing the resulting vectors, or vice versa—does not yield UNTFs, as these two operations do not commute. Nevertheless, by using one of these two techniques, one may assume without loss of generality [2] that either the initial frame is exactly tight and nearly unit norm or, alternatively, that the initial frame is exactly unit norm and nearly tight. The former approach is that taken by [2]: starting with a tight frame that is not unit norm, they solve a differential equation that minimizes *frame energy* while preserving tightness, flowing towards a UNTF; this led to the first genuine solution to the Paulsen problem in the special case where M and N are relatively prime. The latter approach is the one we pursue here.

In particular, starting with a frame that is already unit norm, we try to produce a UNTF from it. Preliminary results to this end were reported in the conference proceedings paper [4]. We accomplish this task by descending against the gradient of the *frame potential*, namely the square of the Hilbert-Schmidt norm of the Gram matrix F^*F, regarded as a function over N copies of the unit sphere $\mathbb{S}_M := \{f \in \mathbb{H}_M : \|f\| = 1\}$:

$$\mathrm{FP} : \mathbb{S}_M^N \to \mathbb{R}, \quad \mathrm{FP}(F) = \|F^*F\|_{\mathrm{HS}}^2 = \sum_{n=1}^{N}\sum_{n'=1}^{N}|\langle f_n, f_{n'}\rangle|^2.$$

Introduced in [1], the frame potential is the total potential energy contained within a given collection of points on the sphere under the action of a *frame force* which encourages orthogonality. As discussed in the next section, one can show that $\mathrm{FP}(F) = \frac{N^2}{M} + \|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}^2$ for any $F \in \mathbb{S}_M^N$. That is, the frame potential is bounded below by $\frac{N^2}{M}$, with

equality if and only if F is a UNTF. The main result of [1] gives that even *local* minimizers of FP are UNTFs. As such, even if no explicit constructions of such frames were known, they must exist: FP is a continuous function over the compact set $\mathbb{S}_M^N$, and as such, possesses a global minimizer, which is necessarily a local minimizer, which is necessarily a UNTF. This existence argument has been generalized to numerous other settings [3, 5, 11, 15, 18, 19, 20]. Moreover, this fact implies that every local minimizer of FP is necessarily a global minimizer, which is a nice property to have when performing gradient descent; even here, this task is nontrivial however, as there are nonoptimal arrangements at which the first derivative of the frame potential vanishes [1].

The novelty and significance of our work is best gauged by contrasting it with the current state-of-the-art of the Paulsen problem: the technique of [2]. Both approaches give valid solutions to the Paulsen problem and have certain applications for which they are preferable to the other. Instead of assuming our frame is already tight and seeking to become increasingly unit norm [2], we assume we are already unit norm and seek tightness. Rather than needing to solve a differential equation [2], we have an iterative, gradient-descent-based algorithm; our approach only becomes a differential equation when the step size is forced arbitrarily small. While the relative primeness of M and N is an important consideration in both methods, the technique of [2] is only guaranteed to converge in this case, while our convergence argument generalizes to the non-relatively-prime case, albeit in a weaker form. Also, as shown below, our method preserves the group structure of certain UNTF constructions, such as Gabor frames and filter banks, whereas [2] does not.

In the next section, we introduce the fundamental concepts needed to compute the gradient of the frame potential (Theorem 2) and study its group invariance properties (Proposition 3). In Section 3, we find sufficient conditions that guarantee that gradient descent of the frame potential converges to a UNTF at a linear rate (Theorem 6). In the fourth and final section, we show that these sufficient conditions are indeed met provided M and N are relatively prime and the initial frame is already sufficient tight, yielding an answer to the Paulsen problem in this case (Corollary 8). We further discuss how these arguments generalize to the non-relatively-prime case (Theorem 11).

2. The gradient of the frame potential

In this section, we lay the groundwork for our approach to modify a given unit norm frame so as to decrease its distance from tightness. As such, our first priority is to formally define this distance. Let $\{\lambda_m\}_{m=1}^M$ be the eigenvalues of the frame operator FF^* of some unit norm sequence $F = \{f_n\}_{n=1}^N$. Note that since

$$\sum_{m=1}^M \lambda_m = \mathrm{Tr}(FF^*) = \mathrm{Tr}(F^*F) = \sum_{n=1}^N \|f_n\|^2 = N,$$

the average value of these eigenvalues is $\frac{N}{M}$. Moreover, F is a UNTF if and only if $FF^* = \frac{N}{M}\mathrm{I}$, that is, if and only if all the λ_m's are equal to $\frac{M}{N}$. As such, in the past, the *distance from tightness* of a unit norm frame F has usually been defined as $\max_m |\lambda_m - \frac{N}{M}|$. However, as there is no closed-form expression for eigenvalues exist, we propose an alternative measure of tightness, namely the 2-norm of the values $\{\lambda_m - \frac{N}{M}\}_{m=1}^M$:

$$\sum_{m=1}^M (\lambda_m - \tfrac{N}{M})^2 = \left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2 = \mathrm{Tr}[(FF^*)^2] - 2\tfrac{N}{M}\mathrm{Tr}(FF^*) + \tfrac{N^2}{M^2}\mathrm{Tr}(\mathrm{I}) = \mathrm{FP}(F) - \tfrac{N^2}{M}. \tag{2}$$

In particular, we see that $\mathrm{FP}(F) \geq \frac{N^2}{M}$, with equality if and only if F is a UNTF. It therefore makes sense to define our notion of the *distance from tightness* of F to be the easily computable quantity $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}} = (\mathrm{FP}(F) - \frac{N^2}{M})^{\frac{1}{2}}$. Written in this language, the version of the Paulsen problem on which we focus is the following:

> *Given positive integers M and N, find possibly (M, N)-dependent constants δ, C and α such that given any unit norm sequence F such that $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}} \leq \delta$, there necessarily exists a UNTF $\tilde{F}$ such that*
>
> $$\|\tilde{F} - F\|_{\mathrm{HS}} \leq C\left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^{\alpha}. \tag{3}$$

One way to get a ballpark estimate on what these parameters δ, C and α should be, under the best possible circumstances, is to solve a weaker problem: given a unit norm frame F, find $\tilde{F}$ such that $\tilde{F}\tilde{F}^* = \frac{N}{M}\mathrm{I}$ and such that $\|\tilde{F} - F\|_{\mathrm{HS}}$ is minimized; here, we do not require that $\tilde{F}$ be unit norm. Similar problems have been extensively studied in the past—see [2] for references. In brief, we have that for any such $\tilde{F}$ and F, $\|\tilde{F} - F\|_{\mathrm{HS}}^2 = 2N - 2\mathrm{ReTr}(\tilde{F}^*F)$. Taking the singular value decomposition $F = U\Sigma V$ and letting $\tilde{\Sigma} = U^*\tilde{F}V^*$ so that $\tilde{F} = U\tilde{\Sigma}V$, we are therefore seeking to maximize $\mathrm{ReTr}(\tilde{F}^*F) = \mathrm{ReTr}(\tilde{\Sigma}^*\Sigma)$ subject to the restriction that $\tilde{\Sigma}\tilde{\Sigma}^* = \frac{N}{M}\mathrm{I}$. As Σ is "diagonal," this maximum is achieved by letting $\tilde{\Sigma}$ also be "diagonal" with entries $(\frac{N}{M})^{\frac{1}{2}}$, implying

$$\|\tilde{F} - F\|_{\mathrm{HS}}^2 = 2N - 2\mathrm{ReTr}(\tilde{\Sigma}^*\Sigma) \geq 2N - 2(\tfrac{N}{M})^{\frac{1}{2}} \sum_{m=1}^{M} \lambda_m^{\frac{1}{2}} = \sum_{m=1}^{M} \left[\lambda_m^{\frac{1}{2}} - (\tfrac{N}{M})^{\frac{1}{2}}\right]^2.$$

Multiplying the terms in these summands by their conjugates $\lambda_m^{\frac{1}{2}} + (\frac{N}{M})^{\frac{1}{2}}$ then yields

$$\|\tilde{F} - F\|_{\mathrm{HS}}^2 \geq \sum_{m=1}^{M} \frac{\left(\lambda_m - \frac{N}{M}\right)^2}{\left[\lambda_m^{\frac{1}{2}} + (\frac{N}{M})^{\frac{1}{2}}\right]^2} \geq \tfrac{M}{N} \sum_{m=1}^{M} (\lambda_m - \tfrac{N}{M})^2 = \tfrac{M}{N} \left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2.$$

To summarize, the UNTF $\tilde{F}$ which is closest to F necessarily satisfies $\|\tilde{F} - F\|_{\mathrm{HS}} \geq (\frac{M}{N})^{\frac{1}{2}} \|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}$. As such, in our version of the Paulsen problem (3), the best α we should expect is $\alpha = 1$. Indeed, in the case where M and N are relatively prime, we show that $\alpha = 1$ is achievable, provided δ and C are suitably chosen. Meanwhile, when M and N have a common divisor, a simple example, given in Section 4, shows that the best one can expect is $\alpha = \frac{1}{2}$. As we shall see, the key issue with the non-relatively-prime case is that there exist UNTFs which can be partitioned into mutually orthogonal subcollections; at such frames, the geometric structure of the set of surrounding UNTFs is extremely complicated [10].

2.1. The gradient of the frame potential

Now that we have formally defined the distance from tightness of a unit norm frame F to be $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}$, and having further posed the problem we are trying to solve with (3), we turn to our specific approach: a gradient descent of the squared distance from tightness, which, since $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}^2 = \mathrm{FP}(F) - \frac{N^2}{M}$, reduces to a gradient descent of the frame potential. Here, as the domain of optimization $\mathbb{S}_M^N$ is a product of spheres as opposed to the entire space $\mathbb{H}_M^N$, this version of gradient descent differs from the one most commonly used. In particular, given $F = \{f_n\}_{n=1}^N$ in $\mathbb{S}_M^N$ and $G = \{g_n\}_{n=1}^N$ in $\oplus_{n=1}^N f_n^\perp := \{\{g_n\}_{n=1}^N \in \mathbb{H}_M^N : \langle f_n, g_n\rangle = 0,\ \forall n\}$, we use Lemma 2 of [3] along with Taylor's theorem to estimate the change in frame potential as each f_n is pushed along a great circle with tangent velocity g_n:

Proposition 1. *For any $F = \{f_n\}_{n=1}^N \in \mathbb{S}_M^N$ and $G = \{g_n\}_{n=1}^N \in \oplus_{n=1}^N f_n^\perp$, let $f_n(t) := \cos(\|g_n\|t) f_n - \sin(\|g_n\|t)\frac{g_n}{\|g_n\|}$ whenever $g_n \neq 0$, and let $f_n(t) := f_n$ otherwise. Then, $F(t) = \{f_n(t)\}_{n=1}^N \in \mathbb{S}_M^N$ for any $t \in \mathbb{R}$ and satsifies*

$$\|F(t) - F\|_{\mathrm{HS}}^2 \leq t^2 \sum_{n=1}^{N} \|g_n\|^2, \tag{4}$$

$$\mathrm{FP}(F(t)) \leq \mathrm{FP}(F) - 4t\mathrm{Re} \sum_{n=1}^{N} \langle FF^* f_n, g_n\rangle + 8Nt^2 \sum_{n=1}^{N} \|g_n\|^2. \tag{5}$$

Proof. It is straightforward to show that $\|f_n(t)\| = 1$ for all $n = 1, \ldots, N$ and all $t \in \mathbb{R}$. To show (4), note that for any n such that $g_n \neq 0$, we have

$$\|f_n(t) - f_n\|^2 = (\cos(\|g_n\|t) - 1)^2 + \sin^2(\|g_n\|t) = 4\sin^2(\|g_n\|t/2) \leq \|g_n\|^2 t^2. \tag{6}$$

As (6) also immediately holds for any n such that $g_n = 0$, we may sum (6) over all n to conclude (4). To prove (5), we apply Taylor's theorem to $\varphi(t) = \mathrm{FP}(F(t))$ at $t = 0$:

$$\varphi(t) \leq \varphi(0) + t\dot{\varphi}(0) + \tfrac{1}{2}t^2 \max_{s\in\mathbb{R}} |\ddot{\varphi}(s)|. \tag{7}$$

To compute the terms in (7), note that $\dot{f}_n(t) = -\|g_n\|\sin(\|g_n\|t)f_n - \cos(\|g_n\|t)g_n$ for any n such that $g_n \neq 0$, a fact that also holds trivially when $g_n = 0$, since $f_n(t)$ is constant. In particular, $\dot{f}_n(0) = -g_n$ for all $n = 1, \ldots, N$. The expression for $\dot{\varphi}(t)$ given in Lemma 2 of [3] then gives

$$\dot{\varphi}(0) = 4\mathrm{ReTr}(\dot{F}^*(0)F(0)F^*(0)F(0)) = 4\mathrm{ReTr}(-G^*FF^*F) = -4\mathrm{Re}\sum_{n=1}^{N}\langle G^*FF^*Fe_n, e_n\rangle = -4\mathrm{Re}\sum_{n=1}^{N}\langle FF^*f_n, g_n\rangle, \quad (8)$$

where $\{e_n\}_{n=1}^N$ is the standard basis of $\mathbb{H}_N$. Next, as $\ddot{f}_n(t) = -\|g_n\|^2 f_n(t)$ for any n, we further have

$$\mathrm{Tr}(\ddot{F}^*(t)F(t)F^*(t)F(t)) = \sum_{n=1}^{N}\langle \ddot{F}^*(t)F(t)F^*(t)F(t)e_n, e_n\rangle = \sum_{n=1}^{N}\langle F^*(t)f_n(t), F^*(t)\ddot{f}_n(t)\rangle = -\sum_{n=1}^{N}\|g_n\|^2\|F^*(t)f_n(t)\|^2. \quad (9)$$

Substituting (9) into the expression for $\ddot{\varphi}(t)$ given in Lemma 2 of [3] yields

$$\ddot{\varphi}(t) = -4\sum_{n=1}^{N}\|g_n\|^2\|F^*(t)f_n(t)\|^2 + 4\|\dot{F}^*(t)F(t)\|_{\mathrm{HS}}^2 + 2\|\dot{F}(t)F^*(t) + F(t)\dot{F}^*(t)\|_{\mathrm{HS}}^2. \quad (10)$$

To bound (10), note that $\|F(t)\|_{\mathrm{HS}}^2 = \sum_{n=1}^N \|f_n(t)\|^2 = N$ and $\|\dot{F}(t)\|_{\mathrm{HS}}^2 = \sum_{n=1}^N \|\dot{f}_n(t)\|^2 = \sum_{n=1}^N \|g_n\|^2$, and thus

$$\begin{aligned}
|\ddot{\varphi}(t)| &\leq 4\sum_{n=1}^{N}\|g_n\|^2\|F^*(t)f_n(t)\|^2 + 4\|\dot{F}^*(t)F(t)\|_{\mathrm{HS}}^2 + 2\|\dot{F}(t)F^*(t) + F(t)\dot{F}^*(t)\|_{\mathrm{HS}}^2 \\
&\leq 4\sum_{n=1}^{N}\|g_n\|^2\|F(t)\|_2^2\|f_n(t)\|^2 + 4\|\dot{F}^*(t)F(t)\|_{\mathrm{HS}}^2 + 2\Big(\|\dot{F}(t)F^*(t)\|_{\mathrm{HS}} + \|F(t)\dot{F}^*(t)\|_{\mathrm{HS}}\Big)^2 \\
&\leq 4\sum_{n=1}^{N}\|g_n\|^2\|F(t)\|_{\mathrm{HS}}^2 + 12\|\dot{F}(t)\|_{\mathrm{HS}}^2\|F(t)\|_{\mathrm{HS}}^2 \\
&= 16N\sum_{n=1}^{N}\|g_n\|^2. \quad (11)
\end{aligned}$$

Substituting (8) and (11) into (7) yields (5). □

Considering the Taylor expansion of $\mathrm{FP}(F(t))$ given in (4), one might expect the *gradient* of FP over $\mathbb{S}_M^N$, namely the choice of vectors $\{g_n\}_{n=1}^N$, modulo positive scalar multiples, which maximizes the linear term $\mathrm{Re}\sum_{n=1}^N\langle FF^*f_n, g_n\rangle$, to be given by $g_n = FF^*f_n$ for all $n = 1, \ldots, N$. Indeed, one may show that this would be the correct gradient if we regarded the frame potential as a functional over the entire space $\mathbb{H}_M^N$. However, since we are optimizing over $\mathbb{S}_M^N$, we require that $\{g_n\}_{n=1}^N \in \oplus_{n=1}^N f_n^\perp$. Therefore, we instead take $\{g_n\}_{n=1}^N$ to be the projection of $\{FF^*f_n\}_{n=1}^N$ onto $\oplus_{n=1}^N f_n^\perp$. In the next result, we formally verify that such a choice is optimal.

Theorem 2. *Pick $F = \{f_n\}_{n=1}^N \in \mathbb{S}_M^N$, and let P_n denote the orthogonal projection from $\mathbb{H}_M$ onto the orthogonal complement of f_n. Then, the minimizer of the bound in* (5) *over all $t \in \mathbb{R}$ and $\{g_n\}_{n=1}^N \in \oplus_{n=1}^N f_n^\perp$ is given by $t = \frac{1}{4N}$ and*

$$g_n = P_nFF^*f_n = FF^*f_n - \langle FF^*f_n, f_n\rangle f_n, \quad n = 1, \ldots, N. \quad (12)$$

Moreover, for any $t \in \mathbb{R}$, this choice for $\{g_n\}_{n=1}^N$ gives

$$\|F(t) - F\|_{\mathrm{HS}}^2 \leq t^2\sum_{n=1}^{N}\|P_nFF^*f_n\|^2, \quad (13)$$

$$\mathrm{FP}(F(t)) \leq \mathrm{FP}(F) - 4t(1 - 2Nt)\sum_{n=1}^{N}\|P_nFF^*f_n\|^2. \quad (14)$$

Proof. We seek to minimize

$$-4t\mathrm{Re}\sum_{n=1}^{N}\langle FF^*f_n, g_n\rangle + 8Nt^2\sum_{n=1}^{N}\|g_n\|^2 = \tfrac{2}{N}\sum_{n=1}^{N}\mathrm{Re}\langle -FF^*f_n + 2Ntg_n, 2Ntg_n\rangle \tag{15}$$

over all $\{g_n\}_{n=1}^N \in \mathbb{S}_M^N$ and all $t \in \mathbb{R}$. We note immediately from (15) that the optimal $\{g_n\}_{n=1}^N$ and t are not unique, though we now show that their product is. Indeed, we have $P_n g_n = g_n$, and therefore

$$\begin{aligned}\mathrm{Re}\langle -FF^*f_n + 2Ntg_n, 2Ntg_n\rangle &= \mathrm{Re}\langle -FF^*f_n + 2Ntg_n, 2NtP_ng_n\rangle\\ &= \mathrm{Re}\langle -P_nFF^*f_n + 2Ntg_n, 2Ntg_n\rangle\\ &= \tfrac{1}{4}(\|-P_nFF^*f_n + 4Ntg_n\|^2 - \|-P_nFF^*f_n\|^2)\\ &\geq -\tfrac{1}{4}\|P_nFF^*f_n\|^2,\end{aligned}$$

with equality if and only if $-P_nFF^*f_n + 4Ntg_n = 0$. Thus, to minimize (15), and consequently to minimize the upper bound in (5), we may take $t = \frac{1}{4N}$ and $g_n = P_nFF^*f_n$, as claimed. Moreover, substituting these choices of g_n's into (4) and (5) yields (13) and (5), respectively. □

Note that for any $t \in (0, \frac{1}{2N})$, Theorem 2 prescribes a direction and step size to travel from a given $F \in \mathbb{S}_M^N$ which guarantees a predictable decrease in frame potential. Throughout the remainder of this paper, we fix any such t and repeatedly apply Theorem 2 to produce a sequence of iterations which, in many cases, is guaranteed to converge to a UNTF. One may also consider what happens to this sequence of iterations as t is taken ever smaller; as $t \to 0$, we expect to approach a solution to the system of nonlinear ordinary differential equations:

$$\dot{f}_n(s) = -\Big(F(s)F^*(s)f_n(s) - \Big\langle F(s)F^*(s)f_n(s),\, f_n(s)\Big\rangle f_n(s)f_n(s)\Big), \quad \forall n = 1,\dots,N,$$

a matter we leave for future research.

2.2. The preservation of group structure

Many popular examples of unit norm frames, such as oversampled filter banks and Gabor frames, have a group structure. In particular, such frames are the *orbit* $\{U_if_j\}_{i\in\mathcal{I},j\in\mathcal{J}}$ of a collection of unit vectors $\{f_j\}_{j\in\mathcal{J}}$ under the action of a collection of unitary operators $\{U_i\}_{i\in\mathcal{I}}$. While such frames inherently consist of unit norm vectors, it can be difficult to ensure their tightness [9, 11]. As such, it would be valuable to have a technique which increases the tightness of such frames without sacrificing their group structure. The next result shows that the technique of Theorem 2 does precisely this, provided the unitary operators are known to commute with the frame operator.

Proposition 3. *Let the orbit $F = \{f_{i,j}\}_{i\in\mathcal{I},j\in\mathcal{J}} = \{U_if_j\}_{i\in\mathcal{I},j\in\mathcal{J}}$ of unit vectors have the property that every unitary matrix U_i commutes with its frame operator FF^*. Then, pushing these vectors along the tangent directions $\{g_{i,j}\}_{i\in\mathcal{I},j\in\mathcal{J}}$ given in* (12) *produces new collections of vectors which possess this same group structure: $F(t) = \{U_if_j(t)\}_{i\in\mathcal{I},j\in\mathcal{J}}$.*

Proof. We have $f_{i,j}(t) = \cos(\|g_{i,j}\|t)f_{i,j} - \sin(\|g_{i,j}\|t)\frac{g_{i,j}}{\|g_{i,j}\|}$ where $g_{i,j} := P_{i,j}FF^*f_{i,j}$. That is,

$$g_{i,j} = FF^*U_if_j - \Big\langle FF^*U_if_j, U_if_j\Big\rangle U_if_j = U_iFF^*f_j - \Big\langle U_iFF^*f_j, U_if_j\Big\rangle U_if_j = U_i\Big(FF^*f_j - \Big\langle FF^*f_j, f_j\Big\rangle f_j\Big) = U_ig_j,$$

where $g_j := FF^*f_j - \langle FF^*f_j, f_j\rangle f_j$. We thus have that $f_{i,j}(t) = U_jf_i(t)$, as claimed:

$$f_{i,j}(t) = \cos(\|U_ig_j\|t)U_if_j - \sin(\|U_ig_j\|t)\tfrac{U_ig_j}{\|U_ig_j\|} = U_i\Big(\cos(\|g_j\|t)f_j - \sin(\|g_j\|t)\tfrac{g_j}{\|g_j\|}\Big) = U_if_j(t).$$ □

For example, consider the space of discrete M-periodic signals $\ell(\mathbb{Z}_M) = \{f : \mathbb{Z} \to \mathbb{C} : f(m+M) = f(m),\ \forall m\}$. Letting $M = AC$, the *synthesis filter bank* associated with some unit norm vectors $\{f_j\}_{j\in\mathcal{J}}$ is $\{\mathrm{T}^{Ai}f_j\}_{i=0,\,j\in\mathcal{J}}^{C-1}$, where T is the *translation* operator $(\mathrm{T}f)(m) := f(m-1)$. As one may verify that $FF^*\mathrm{T}^{Ai} = \mathrm{T}^{Ai}FF^*$, Proposition 3 guarantees that evolving the f_j's according to Theorem 2 preserves this filter bank structure. Letting $M = BD$, one can further consider the Gabor subclass of filter bank frames: the *Gabor system* associated with some unit norm f is $\{\mathrm{T}^{Ai}\mathrm{E}^{Bj}f\}_{i=0,\,j=0}^{C-1,D-1}$, where E is the *modulation* operator $(\mathrm{E}f)(m) = \mathrm{e}^{\frac{2\pi im}{M}}f(m)$. Though the operators E and T do not

commute, we nevertheless have that $\mathrm{ET} = \mathrm{e}^{\frac{2\pi i}{M}}\mathrm{TE}$, a fact which suffices to guarantee that $FF^*\mathrm{T}^{Ai}\mathrm{E}^{Bj} = \mathrm{T}^{Ai}\mathrm{E}^{Bj}FF^*$, and so Proposition 3 guarantees that the method of Theorem 2 preserves the Gabor structure. In particular, one need only evolve f itself, rather than the entirety of its modulates and translates. That is, one need only compute

$$FF^*f = \sum_{i=0}^{C-1}\sum_{j=0}^{D-1}\langle f, \mathrm{T}^{ai}\mathrm{E}^{bj}f\rangle \mathrm{T}^{ai}\mathrm{E}^{bj}f$$

and consider $f(t) = \cos(\|g\|t)f - \sin(\|g\|t)\frac{g}{\|g\|}$, where $g = FF^*f - \langle FF^*f, f\rangle f$ and $t \in (0, \frac{1}{2N})$. By iteratively applying this procedure, one produces Gabor frames of ever-increasing tightness.

3. Sufficient conditions for linear convergence of gradient descent

We now take a given unit norm sequence $F_0 := F = \{f_n\}_{n=1}^N$, and iteratively apply the main result of the previous section—Theorem 2—to produce a sequence $\{F_k\}_{k=0}^\infty$ of unit norm sequences of increasing tightness. To be clear, fixing any $t \in (0, \frac{1}{2N})$, and given any unit norm sequence $F_k = \{f_n^{(k)}\}_{n=1}^N$, we first compute $G_k = \{g_n^{(k)}\}_{n=1}^N$:

$$g_n^{(k)} = P_n^{(k)}F_kF_kf_n^{(k)} = F_kF_kf_n^{(k)} - \langle F_kF_kf_n^{(k)}, f_n^{(k)}\rangle f_n^{(k)}, \quad \forall n = 1, \ldots, N. \tag{16}$$

We then define $F_k = \{f_n^{(k+1)}\}_{n=1}^N$ as follows:

$$f_n^{(k+1)} := \begin{cases} \cos(\|g_n^{(k)}\|t)f_n^{(k)} - \sin(\|g_n^{(k)}\|t)\frac{g_n^{(k)}}{\|g_n^{(k)}\|}, & g_n^{(k)} \neq 0, \\ f_n^{(k)}, & g_n^{(k)} = 0. \end{cases} \tag{17}$$

While Theorem 2 guarantees that the values of $\|F_kF_k^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}$ are decreasing, it does not guarantee that this decrease is strict, nor that it decreases to zero in the limit, nor that the F_k's themselves converge. Indeed, gradient descent of the frame potential does not necessarily converge to a UNTF: despite the fact that every local minimizer of the frame potential is also a global minimizer, there do exist suboptimal *critical frames* F at which the gradient G vanishes [1]. In this section, we provide conditions which suffice to avoid such nonoptimal critical frames, and moreover, guarantee that the iterative application of (16) and (17) produces a sequence of unit norm frames which indeed converges to a UNTF $F_\infty = \lim_k F_k$ that is close to $F = F_0$. To do this, note that a unit norm sequence F is critical with respect to the frame potential if and only if its gradient G vanishes, which occurs precisely when each f_n is an eigenvector of the frame operator FF^*. As noted in [1], this occurs precisely when F can be partitioned into a collection of subsequences, each of which is a unit norm tight frame for its span. Here, the key is to recognize that in this setting, such orthogonality is actually one's enemy. To be precise, we make the following definition:

Definition 4. A sequence $\{f_n\}_{n=1}^N \in \mathbb{S}_M^N$ is termed *orthogonally partitionable (OP)* if there exists a nontrivial partition $\mathcal{I} \sqcup \mathcal{J} = \{1, \ldots, N\}$ such that $|\langle f_i, f_j\rangle| = 0$ for every $i \in \mathcal{I}$, $j \in \mathcal{J}$. More generally, it is *ε-orthogonally partitionable (ε-OP)* if there exists a nontrivial partition $\mathcal{I} \sqcup \mathcal{J} = \{1, \ldots, N\}$ such that $|\langle f_i, f_j\rangle| < \varepsilon$ for every $i \in \mathcal{I}$, $j \in \mathcal{J}$.

Thus, one way to ensure $G \neq 0$ is to have that F is not OP. Indeed, as we show in the following result, if F is not ε-OP, then the amount F's frame potential decreases in one iteration of gradient descent, as given in Theorem 2, is at least some fixed percentage of F's distance from tightness.

Theorem 5. *Let $\varepsilon \in (0, \frac{1}{\mathrm{j}}$, and take $F \in \mathbb{S}_M^N$ satisfying $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}} \leq \frac{N}{2M}$. Let P_n denote the orthogonal projection from $\mathbb{H}_M$ onto the orthogonal complement of f_n. If F is not ε-orthogonally partitionable, then*

$$\tfrac{\varepsilon^2}{4M^4}\left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2 \leq \sum_{n=1}^N \|P_nFF^*f_n\|^2 \leq 4N\left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2. \tag{18}$$

Proof. Let $\{\lambda_m\}_{m=1}^M$ denote the eigenvalues of FF^*, arranged in increasing order, with corresponding orthonormal eigenbasis $\{e_m\}_{m=1}^M$. Decomposing any f_n in terms of this eigenbasis gives

$$\gamma_n := \langle FF^*f_n, f_n\rangle = \left\langle FF^*\sum_{m=1}^M \langle f_n, e_m\rangle e_m, f_n\right\rangle = \sum_{m=1}^M \lambda_m|\langle f_n, e_m\rangle|^2.$$

That is, each γ_n is a convex combination of FF^*'s spectrum. Since, as noted previously, $\frac{N}{M}$ is the average of the λ_m's, we therefore have $\gamma_n, \frac{N}{M} \in [\lambda_1, \lambda_M]$, and so for any m and n,

$$(\lambda_m - \gamma_n)^2 \leq (\lambda_M - \lambda_1)^2 \leq 4 \max_{m'} (\lambda_{m'} - \tfrac{N}{M})^2 \leq 4 \sum_{m'=1}^{M} (\lambda_{m'} - \tfrac{N}{M})^2 = 4 \left\| FF^* - \tfrac{N}{M} \mathrm{I} \right\|_{\mathrm{HS}}^2. \tag{19}$$

Also, by the definitions of P_n and γ_n, we have $\sum_{n=1}^N \|P_n FF^* f_n\|^2 = \sum_{n=1}^N \|(FF^* - \gamma_n \mathrm{I}) f_n\|^2$. Decomposing each f_n in terms of the e_m's therefore gives

$$\sum_{n=1}^{N} \|P_n FF^* f_n\|^2 = \sum_{n=1}^{N} \left\| (FF^* - \gamma_n \mathrm{I}) \sum_{m=1}^{M} \langle f_n, e_m \rangle e_m \right\|^2 = \sum_{n=1}^{N} \left\| \sum_{m=1}^{M} (\lambda_m - \gamma_n) \langle f_n, e_m \rangle e_m \right\|^2 = \sum_{n=1}^{N} \sum_{m=1}^{M} (\lambda_m - \gamma_n)^2 |\langle f_n, e_m \rangle|^2. \tag{20}$$

From here, we apply (19) to get the right-hand inequality of (18):

$$\sum_{n=1}^{N} \|P_n FF^* f_n\|^2 \leq 4 \left\| FF^* - \tfrac{N}{M} \mathrm{I} \right\|_{\mathrm{HS}}^2 \sum_{n=1}^{N} \sum_{m=1}^{M} |\langle f_n, e_m \rangle|^2 = 4N \left\| FF^* - \tfrac{N}{M} \mathrm{I} \right\|_{\mathrm{HS}}^2.$$

Note that this inequality holds in general, that is, for any $F \in \mathbb{S}_M^N$. We now seek the left-hand inequality of (18). Since the largest gap between successive eigenvalues is no smaller than the average gap, there necessarily exists an m_0 that satisfies

$$\lambda_{m_0+1} - \lambda_{m_0} \geq \tfrac{1}{M-1} (\lambda_M - \lambda_1) \geq \tfrac{1}{M} (\lambda_M - \lambda_1). \tag{21}$$

Define $\mathcal{I} := \{n : \gamma_n < \frac{1}{2}(\lambda_{m_0} + \lambda_{m_0+1})\}$, $\mathcal{J} := \{1, \ldots, N\} \setminus \mathcal{I}$. This partitions the γ_n's according to where they lie in relation to the midpoint $\frac{1}{2}(\lambda_{m_0} + \lambda_{m_0+1})$ of the largest gap between eigenvalues. Therefore, the λ_m's lying above this midpoint are at least half the gap away, namely at least $\frac{1}{2}(\lambda_{m_0+1} - \lambda_{m_0}) \geq \frac{1}{2M}(\lambda_M - \lambda_1)$ away, from the γ_n's lying below the midpoint, and vice versa. In fact, when $m \geq m_0 + 1$ and $n \in \mathcal{I}$, or when $m \leq m_0$ and $n \in \mathcal{J}$, we have

$$(\lambda_m - \gamma_n)^2 \geq \left[\tfrac{1}{2M} (\lambda_M - \lambda_1) \right]^2 \geq \tfrac{1}{4M^2} \max_m (\lambda_m - \tfrac{N}{M})^2 \geq \tfrac{1}{4M^3} \sum_m (\lambda_m - \tfrac{N}{M})^2 = \tfrac{1}{4M^3} \left\| FF^* - \tfrac{N}{M} \mathrm{I} \right\|_{\mathrm{HS}}^2. \tag{22}$$

That said, if $i \in \mathcal{I}$ and $j \in \mathcal{J}$, then regardless of m, λ_m is on one side of the midpoint $\frac{1}{2}(\lambda_{m_0} + \lambda_{m_0+1})$, and either γ_i or γ_j is on the other side, implying

$$\max \left\{ (\lambda_m - \gamma_i)^2, (\lambda_m - \gamma_j)^2 \right\} \geq \tfrac{1}{4M^3} \left\| FF^* - \tfrac{N}{M} \mathrm{I} \right\|_{\mathrm{HS}}^2. \tag{23}$$

Now suppose both $\mathcal{I}$ and $\mathcal{J}$ are nonempty. Since F is not ε-OP, there exists $i \in \mathcal{I}$ and $j \in \mathcal{J}$ such that $\varepsilon \leq |\langle f_i, f_j \rangle|$. Decomposing over the eigenbasis, we therefore have

$$\varepsilon^2 \leq |\langle f_i, f_j \rangle|^2 \leq \left(\sum_{m=1}^{M} |\langle f_i, e_m \rangle| |\langle f_j, e_m \rangle| \right)^2 \leq M \sum_{m=1}^{M} |\langle f_i, e_m \rangle|^2 |\langle f_j, e_m \rangle|^2 \leq M \sum_{m=1}^{M} \min \left\{ |\langle f_i, e_m \rangle|^2, |\langle f_j, e_m \rangle|^2 \right\}, \tag{24}$$

where the last inequality uses $|\langle f_n, e_m \rangle| \leq \|f_n\| \|e_m\| = 1$. Recalling (20), we isolate the ith and jth terms:

$$\begin{aligned}
\sum_{n=1}^{N} \|P_n FF^* f_n\|^2 &= \sum_{n=1}^{N} \sum_{m=1}^{M} (\lambda_m - \gamma_n)^2 |\langle f_n, e_m \rangle|^2 \\
&\geq \sum_{m=1}^{M} \left((\lambda_m - \gamma_i)^2 |\langle f_i, e_m \rangle|^2 + (\lambda_m - \gamma_j)^2 |\langle f_j, e_m \rangle|^2 \right) \\
&\geq \sum_{m=1}^{M} \max \left\{ (\lambda_m - \gamma_i)^2, (\lambda_m - \gamma_j)^2 \right\} \min \left\{ |\langle f_i, e_m \rangle|^2, |\langle f_j, e_m \rangle|^2 \right\}.
\end{aligned}$$

From here, we apply (23) and (24) to get

$$\sum_{n=1}^{N}\|P_nFF^*f_n\|^2 \geq \tfrac{1}{4M^3}\left\|FF^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2\sum_{m=1}^{M}\min\left\{|\langle f_i,e_m\rangle|^2,|\langle f_j,e_m\rangle|^2\right\} \geq \tfrac{\varepsilon^2}{4M^4}\left\|FF^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2.$$

Therefore, we indeed have the left-hand inequality of (18) in the case where both $\mathcal{I}$ and $\mathcal{J}$ are nonempty. We now turn to the case where either $\mathcal{I}$ or $\mathcal{J}$ is empty. We have

$$\max_m(\lambda_m-\tfrac{N}{M})^2 \leq \sum_{m=1}^{M}(\lambda_m-\tfrac{N}{M})^2 = \left\|FF^*-\tfrac{N}{M}\right\|_{\mathrm{HS}}^2 \leq (\tfrac{N}{2M})^2, \tag{25}$$

where the last inequality follows from one of our assumptions. Therefore, recalling m_0 from (21), we have

$$\sum_{n=1}^{N}|\langle f_n,e_{m_0}\rangle|^2 = \|F^*e_{m_0}\|^2 = \langle FF^*e_{m_0},e_{m_0}\rangle = \lambda_{m_0} \geq \lambda_1 \geq \tfrac{N}{M}-\max_m\left|\lambda_m-\tfrac{N}{M}\right| \geq \tfrac{N}{2M}, \tag{26}$$

where the last inequality is by (25). In particular, if $\mathcal{I}$ is empty, we recall (20), isolating its m_0th term:

$$\sum_{n=1}^{N}\|P_nFF^*f_n\|^2 = \sum_{n=1}^{N}\sum_{m=1}^{M}(\lambda_m-\gamma_n)^2|\langle f_n,e_m\rangle|^2 \geq \sum_{n=1}^{N}(\lambda_{m_0}-\gamma_n)^2|\langle f_n,e_{m_0}\rangle|^2. \tag{27}$$

Since $\mathcal{I}=\emptyset$, then $\mathcal{J}=\{1,\dots,N\}$, and thus (22) holds for $m=m_0$ and all n. Coupled with (26) and (27), this implies

$$\sum_{n=1}^{N}\|P_nFF^*f_n\|^2 \geq \tfrac{1}{4M^3}\left\|FF^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2\sum_{n=1}^{N}|\langle f_n,e_{m_0}\rangle|^2 \geq \tfrac{N}{8M^4}\left\|FF^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2 \geq \tfrac{\varepsilon^2}{4M^4}\left\|FF^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2,$$

where the last inequality uses $\varepsilon^2\leq 1\leq\frac{N}{2}$. This proves the left-hand inequality of (18) in the case where $\mathcal{I}$ is empty. A similar argument—isolating the (m_0+1)st term in (20)—holds in the remaining case where $\mathcal{J}$ is empty. □

The previous result, along with Theorem 2, guarantees a certain decrease in frame potential, provided the given frame F is not ε-OP. In the next result, we show that if, when performing the gradient descent steps (16) and (17), one can ensure that each iteration F_k is not ε-OP for some $\varepsilon>0$ independent of k, then gradient descent converges to a nearby UNTF at a linear rate.

Theorem 6. *Fix $\varepsilon\in(0,1]$ and $t\in(0,\frac{1}{2N})$, take $F_0=\{f_n^{(0)}\}_{n=1}^N\in\mathbb{S}_M^N$ satisfying $\|F_0F_0^*-\frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}\leq\frac{N}{2M}$, and iterate $F_{k+1}:=F_k(t)$ as in* (16) *and* (17)*. If, for any fixed K, we have that F_k is not ε-orthogonally partitionable for all $k=0,\dots,K-1$, then the Kth iteration F_K satisfies*

$$\|F_K-F_0\|_{\mathrm{HS}} \leq \tfrac{4M^4N^{\frac{1}{2}}}{(1-2Nt)\varepsilon^2}\left\|F_0F_0^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}, \tag{28}$$

$$\left\|F_KF_K^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}} \leq \left(1-\tfrac{t(1-2Nt)\varepsilon^2}{M^4}\right)^{\frac{K}{2}}\left\|F_0F_0^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}. \tag{29}$$

Moreover, if F_k is not ε-orthogonally partitionable for any k, then $F_\infty:=\lim_k F_k$ exists and is a unit norm tight frame within (28) *from F_0.*

Proof. Define $\gamma:=\frac{\varepsilon^2}{4M^4}$, and suppose F_k is not ε-OP for $k=0,\dots,K-1$. Then combining (2), (14) and the lower bound in (18) gives that $F_{k+1}:=F_k(t)$ satisfies

$$\begin{aligned}\left\|F_{k+1}F_{k+1}^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2 &= \mathrm{FP}(F_k(t))-\tfrac{N^2}{M}\\ &\leq \mathrm{FP}(F_k)-\tfrac{N^2}{M}-4t(1-2Nt)\sum_{n=1}^{N}\|P_n^{(k)}F_kF_k^*f_n^{(k)}\|^2\\ &\leq [1-4t(1-2Nt)\gamma]\left\|F_kF_k^*-\tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2.\end{aligned}$$

From here, one may proceed inductively to find that

$$\left\|F_kF_k^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2 \leq [1 - 4t(1-2Nt)\gamma]^k \left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2, \tag{30}$$

which proves (29), recalling $\gamma := \frac{\varepsilon^2}{4M^4}$. Next, let $\delta := 4N$. To prove (28), we use (13), the upper bound in (18), and (30) to obtain

$$\|F_{k+1} - F_k\|_{\mathrm{HS}}^2 \leq t^2 \sum_{n=1}^{N} \|P_n^{(k)} F_k F_k^* f_n^{(k)}\|^2 \leq t^2\delta \left\|F_kF_k^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2 \leq t^2\delta[1 - 4t(1-2Nt)\gamma]^k \left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}^2 \tag{31}$$

for all $k = 0, \ldots, K-1$. In particular, for any $K' < K$, we can bound $\|F_K - F_{K'}\|_{\mathrm{HS}}$ in terms of a geometric series; since $t \in (0, \frac{1}{2N})$ and $\gamma = \frac{\varepsilon^2}{4M^4}$ with $\varepsilon \in (0, 1]$, this series is guaranteed to converge:

$$\|F_K - F_{K'}\|_{\mathrm{HS}} \leq \sum_{k=K'}^{K-1} \|F_{k+1} - F_k\|_{\mathrm{HS}} \leq t\delta^{\frac{1}{2}} \Big(\sum_{k=K'}^{\infty} [1 - 4t(1-2Nt)\gamma]^{\frac{k}{2}} \Big) \left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}. \tag{32}$$

In particular, letting $K' = 0$ in (32) yields (28):

$$\|F_K - F_0\|_{\mathrm{HS}} \leq \left(\frac{t\delta^{\frac{1}{2}}}{1 - [1 - 4t(1-2Nt)\gamma]^{\frac{1}{2}}} \right) \left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}} \leq \frac{\delta^{\frac{1}{2}}}{2(1-2Nt)\gamma} \left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}, \tag{33}$$

where we have used the fact that $(1-x)^{\frac{1}{2}} \leq 1 - \frac{1}{2}x$.

Now suppose F_k is never ε-OP for any k, and so (32) holds for all $K' < K$. In particular, as the series in (32) vanishes (independently of K) as K' grows large, we have that $\{F_k\}_{k=0}^{\infty}$ is a Cauchy sequence. As $\mathbb{S}_M^N$ is complete, $F_\infty := \lim_k F_k$ exists. Taking the limit of (30) yields $\|F_\infty F_\infty^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}} = 0$, and so F_∞ is a UNTF. Meanwhile, taking the limit of (33) yields our final conclusion, namely that F_∞ also satisfies (28):

$$\|F_\infty - F_0\|_{\mathrm{HS}} \leq \frac{\delta^{\frac{1}{2}}}{2(1-2Nt)\gamma} \left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}} = \frac{4M^4N^{\frac{1}{2}}}{(1-2Nt)\varepsilon^2} \left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}. \qquad \square$$

4. Solutions to the Paulsen problem

In the previous section, we applied gradient descent to $F_0 \in \mathbb{S}_M^N$ to produce a sequence of iterates $\{F_k\}_{k=0}^{\infty}$. We showed that if F_0 is sufficiently tight and if all resulting F_k's are not ε-OP for some fixed $\varepsilon > 0$, then this sequence converges to a UNTF at a linear rate. In this section, we show that such an ε always exists, provided M and N are relatively prime. Meanwhile, in the non-relatively-prime case, we give an example that shows such ε's are not guaranteed to exist. In this case, our gradient descent algorithm's rate of convergence is threatened whenever our frame becomes nearly OP; to overcome this threat, we "jump" from our current iterate to a nearby OP frame, and then continue gradient descent on the individual subframes over their respective subspaces. In so doing, we are able to give solutions to the Paulsen problem (3) even in the non-relatively-prime case.

4.1. Case I: M and N are relatively prime

Theorem 6 guarantees that gradient descent converges to a UNTF at a linear rate, provided the iterations never become ε-OP for all arbitrarily small ε's. When M and N are relatively prime, this is not a problem:

Theorem 7. *Take $F \in \mathbb{S}_M^N$ with M and N relatively prime. If $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}^2 \leq \frac{2}{M^3}$, then F is not $(\frac{1}{M^8N^4})$-orthogonally partitionable.*

Proof. We prove by contrapositive: take $F \in \mathbb{S}_M^N$ with M and N relatively prime, and suppose F is ε-OP with $\varepsilon := \frac{1}{M^8N^4}$; we show that $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}^2 > \frac{2}{M^3}$. Since F is ε-OP, there exists a nontrivial partition $\mathcal{I} \sqcup \mathcal{J} = \{1, \ldots, N\}$ such that $|\langle f_i, f_j \rangle| < \varepsilon$ for every $i \in \mathcal{I}$, $j \in \mathcal{J}$. Define $F_{\mathcal{I}} := \{f_i\}_{i \in \mathcal{I}}$ and $F_{\mathcal{J}} := \{f_j\}_{j \in \mathcal{J}}$. The frame operator $F_{\mathcal{I}}F_{\mathcal{I}}^*$ has eigenvalues $\{\lambda_{\mathcal{I},m}\}_{m=1}^{M}$ and eigenvectors $\{e_{\mathcal{I},m}\}_{m=1}^{M}$, and similarly for $F_{\mathcal{J}}F_{\mathcal{J}}^*$. Without loss of generality, we arrange

both sets of eigenvalues in decreasing order. Take $\lambda' := \frac{1}{M^4 N}$, and define $M_{\mathcal{I}} := \#\{m : \lambda_{\mathcal{I},m} \geq \lambda'\}$, and similarly for $M_{\mathcal{J}}$. We know $M_{\mathcal{I}} \geq 1$, since otherwise we have a contradiction:

$$1 \leq |\mathcal{I}| = \mathrm{Tr}(F_{\mathcal{I}}^* F_{\mathcal{I}}) = \mathrm{Tr}(F_{\mathcal{I}} F_{\mathcal{I}}^*) = \sum_{m=1}^{M} \lambda_{\mathcal{I},m} < M\lambda' = \tfrac{1}{M^3 N} < 1.$$

Similarly, $M_{\mathcal{J}} \geq 1$. Moreover, we claim $M_{\mathcal{I}} + M_{\mathcal{J}} \leq M$. Indeed, if not, then $\mathrm{Span}\{e_{\mathcal{I},m}\}_{m=1}^{M_{\mathcal{I}}} \cap \mathrm{Span}\{e_{\mathcal{J},m}\}_{m=1}^{M_{\mathcal{J}}}$ has positive dimension, and so we may find a unit vector u in this subspace. Since $e_{\mathcal{I},m}$ is an eigenvector of $F_{\mathcal{I}} F_{\mathcal{I}}^*$ with eigenvalue $\lambda_{\mathcal{I},m}$, we have

$$u = \sum_{m=1}^{M_{\mathcal{I}}} \langle u, e_{\mathcal{I},m}\rangle e_{\mathcal{I},m} = \sum_{m=1}^{M_{\mathcal{I}}} \langle u, e_{\mathcal{I},m}\rangle \tfrac{1}{\lambda_{\mathcal{I},m}} \sum_{i\in\mathcal{I}} \langle e_{\mathcal{I},m}, f_i\rangle f_i,$$

and we have a similar expression with $\mathcal{J}$. Therefore, we apply the triangle inequality to get

$$\begin{aligned}
1 = |\langle u, u\rangle|^2 &= \Big|\Big\langle \sum_{m=1}^{M_{\mathcal{I}}} \langle u, e_{\mathcal{I},m}\rangle \tfrac{1}{\lambda_{\mathcal{I},m}} \sum_{i\in\mathcal{I}} \langle e_{\mathcal{I},m}, f_i\rangle f_i, \sum_{m=1}^{M_{\mathcal{J}}} \langle u, e_{\mathcal{J},m}\rangle \tfrac{1}{\lambda_{\mathcal{J},m}} \sum_{j\in\mathcal{J}} \langle e_{\mathcal{J},m}, f_j\rangle f_j \Big\rangle\Big| \\
&\leq \sum_{i\in\mathcal{I}} \sum_{m=1}^{M_{\mathcal{I}}} \sum_{j\in\mathcal{J}} \sum_{m'=1}^{M_{\mathcal{J}}} \tfrac{|\langle f_i, f_j\rangle|}{\lambda_{\mathcal{I},m}\lambda_{\mathcal{J},m'}} |\langle u, e_{\mathcal{I},m}\rangle||\langle e_{\mathcal{I},m}, f_i\rangle||\langle u, e_{\mathcal{J},m'}\rangle||\langle e_{\mathcal{J},m'}, f_j\rangle| \\
&\leq \tfrac{\varepsilon}{(\lambda')^2} \sum_{i\in\mathcal{I}} \Big(\sum_{m=1}^{M_{\mathcal{I}}} |\langle u, e_{\mathcal{I},m}\rangle||\langle e_{\mathcal{I},m}, f_i\rangle| \Big) \sum_{j\in\mathcal{J}} \Big(\sum_{m=1}^{M_{\mathcal{J}}} |\langle u, e_{\mathcal{J},m}\rangle||\langle e_{\mathcal{J},m}, f_j\rangle| \Big),
\end{aligned}$$

where the last inequality comes from $|\langle f_i, f_j\rangle| \leq \varepsilon$ and $\lambda_{\mathcal{I},m}, \lambda_{\mathcal{J},m'} \geq \lambda'$. From here, we use $\frac{\varepsilon}{(\lambda')^2} = \frac{1}{N^2}$ and Holder's inequality to get

$$1 \leq \tfrac{1}{N^2} \sum_{i\in\mathcal{I}} \Big(\sum_{m=1}^{M_{\mathcal{I}}} |\langle u, e_{\mathcal{I},m}\rangle|^2 \Big)^{\frac{1}{2}} \Big(\sum_{m=1}^{M_{\mathcal{I}}} |\langle e_{\mathcal{I},m}, f_i\rangle|^2 \Big)^{\frac{1}{2}} \sum_{j\in\mathcal{J}} \Big(\sum_{m=1}^{M_{\mathcal{J}}} |\langle u, e_{\mathcal{J},m}\rangle|^2 \Big)^{\frac{1}{2}} \Big(\sum_{m=1}^{M_{\mathcal{J}}} |\langle e_{\mathcal{J},m}, f_j\rangle|^2 \Big)^{\frac{1}{2}} \leq \tfrac{1}{N^2} |\mathcal{I}||\mathcal{J}| \leq \tfrac{1}{4},$$

a contradiction. As a partial summary, we know $M_{\mathcal{I}}$ and $M_{\mathcal{J}}$ are nonzero and $M_{\mathcal{I}} + M_{\mathcal{J}} \leq M$. Now,

$$|\mathcal{I}| = \mathrm{Tr}(F_{\mathcal{I}}^* F_{\mathcal{I}}) = \mathrm{Tr}(F_{\mathcal{I}} F_{\mathcal{I}}^*) = \sum_{m=1}^{M} \lambda_{\mathcal{I},m} = \sum_{m=1}^{M_{\mathcal{I}}} \lambda_{\mathcal{I},m} + \sum_{m=M_{\mathcal{I}}+1}^{M} \lambda_{\mathcal{I},m},$$

where $\sum_{m=M_{\mathcal{I}}+1}^{M} \lambda_{\mathcal{I},m} < (M - M_{\mathcal{I}})\lambda'$. Therefore, $\sum_{m=1}^{M_{\mathcal{I}}} \lambda_{\mathcal{I},m} > |\mathcal{I}| - (M - M_{\mathcal{I}})\lambda'$, and so Jensen's inequality gives

$$\sum_{m=1}^{M_{\mathcal{I}}} \lambda_{\mathcal{I},m}^2 \geq \tfrac{1}{M_{\mathcal{I}}} \Big(\sum_{m=1}^{M_{\mathcal{I}}} \lambda_{\mathcal{I},m} \Big)^2 > \tfrac{1}{M_{\mathcal{I}}} \Big(|\mathcal{I}| - (M - M_{\mathcal{I}})\lambda' \Big)^2 \geq \tfrac{|\mathcal{I}|^2}{M_{\mathcal{I}}} - \tfrac{2\lambda'|\mathcal{I}|(M-M_{\mathcal{I}})}{M_{\mathcal{I}}}, \tag{34}$$

and similarly for $\mathcal{J}$. We now consider the frame potential of F:

$$\mathrm{FP}(F) = \mathrm{Tr}\big[(FF^*)^2\big] = \mathrm{Tr}\big[(F_{\mathcal{I}} F_{\mathcal{I}}^* + F_{\mathcal{J}} F_{\mathcal{J}}^*)^2\big] = \mathrm{Tr}\big[(F_{\mathcal{I}} F_{\mathcal{I}}^*)^2\big] + \mathrm{Tr}\big[(F_{\mathcal{J}} F_{\mathcal{J}}^*)^2\big] + 2\mathrm{Tr}\big[F_{\mathcal{I}} F_{\mathcal{I}}^* F_{\mathcal{J}} F_{\mathcal{J}}^*\big].$$

Since $\mathrm{Tr}\big[F_{\mathcal{I}} F_{\mathcal{I}}^* F_{\mathcal{J}} F_{\mathcal{J}}^*\big] = \|F_{\mathcal{I}}^* F_{\mathcal{J}}\|_{\mathrm{HS}}^2 \geq 0$, we continue:

$$\mathrm{FP}(F) \geq \sum_{m=1}^{M_{\mathcal{I}}} \lambda_{\mathcal{I},m}^2 + \sum_{m=1}^{M_{\mathcal{J}}} \lambda_{\mathcal{J},m}^2 > \tfrac{|\mathcal{I}|^2}{M_{\mathcal{I}}} + \tfrac{|\mathcal{J}|^2}{M_{\mathcal{J}}} - 2\lambda' \Big[\tfrac{|\mathcal{I}|(M-M_{\mathcal{I}})}{M_{\mathcal{I}}} + \tfrac{|\mathcal{J}|(M-M_{\mathcal{J}})}{M_{\mathcal{J}}} \Big], \tag{35}$$

where the last inequality is by (34). Moreover, considering $M_{\mathcal{I}} + M_{\mathcal{J}} \leq M$, we have

$$\tfrac{|\mathcal{I}|^2}{M_{\mathcal{I}}} + \tfrac{|\mathcal{J}|^2}{M_{\mathcal{J}}} \geq \tfrac{|\mathcal{I}|^2}{M_{\mathcal{I}}} + \tfrac{(N-|\mathcal{I}|)^2}{M-M_{\mathcal{I}}} = \tfrac{N^2}{M} + \tfrac{(|\mathcal{I}|M - M_{\mathcal{I}}N)^2}{MM_{\mathcal{I}}(M-M_{\mathcal{I}})} \geq \tfrac{N^2}{M} + \tfrac{4}{M^3}, \tag{36}$$

where the last inequality uses the fact that M and N are relatively prime—that is, $|\mathcal{I}|M - M_{\mathcal{I}}N$ is a nonzero integer—and $M_{\mathcal{I}}(M - M_{\mathcal{I}}) \leq \frac{M^2}{4}$. Also, since $M_{\mathcal{I}}, M_{\mathcal{J}} \geq 1$, we have

$$\tfrac{|\mathcal{I}|(M-M_{\mathcal{I}})}{M_{\mathcal{I}}} + \tfrac{|\mathcal{J}|(M-M_{\mathcal{J}})}{M_{\mathcal{J}}} \leq (M-1)(|\mathcal{I}| + |\mathcal{J}|) \leq MN. \tag{37}$$

Therefore, combining (35), (36) and (37) gives $\mathrm{FP}(F) > \frac{N^2}{M} + \frac{2}{M^3}$, meaning $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}^2 > \frac{2}{M^3}$. □

Note that Theorem 7 requires sufficient tightness to guarantee that F is not $(\frac{1}{M^8N^4})$-othogonally partitionable. Since gradient descent only decreases the frame potential, Theorem 7 will apply to every subsequent iteration. Therefore, by Theorem 6, gradient descent converges to a UNTF in the relatively prime case:

Corollary 8. *Suppose M and N are relatively prime. Pick $t \in (0, \frac{1}{2N})$, take $F_0 \in \mathbb{S}_M^N$ satisfying $\|F_0F_0^* - \frac{N}{M}\mathrm{I}\|_{\mathrm{HS}}^2 \leq \frac{2}{M^3}$, and iterate $F_{k+1} := F_k(t)$ as in* (16) *and* (17)*. Then, $F_\infty := \lim_k F_k$ exists and is a unit norm tight frame satisfying*

$$\|F_\infty - F_0\|_{\mathrm{HS}} \leq \tfrac{4M^{20}N^{8.5}}{1-2Nt}\left\|F_0F_0^* - \tfrac{N}{M}\mathrm{I}\right\|_{\mathrm{HS}}.$$

This solves the Paulsen problem (3) in the case where M and N are relatively prime. To be explicit, taking $t = \frac{1}{4N}$, we have $\delta = 2^{\frac{1}{2}}M^{-\frac{3}{2}}$, $C = 8M^{20}N^{8.5}$, and $\alpha = 1$. These constants are roughly comparable to those previously given in [2], which were obtained using independent methods. As noted earlier, $\alpha = 1$ is the best one can hope for in any case. In the next subsection, we give an example that shows that these techniques fall apart in the case where M and N share a common divisor, and moreover, that in such cases, we must set our sights lower with respect to α.

4.2. Case II: M and N are not relatively prime

We continue our solution to the Paulsen problem in the remaining case where M and N are not relatively prime. Let's begin this case with an example in two dimensions:

Example 9. Take some real $F \in \mathbb{S}_2^N$, that is, $F = \{(\cos\theta_n, \sin\theta_n)\}_{n=1}^N$ for some collection of θ_n's. In this case, it is known [12] that F is tight precisely when the sum of $\{(\cos 2\theta_n, \sin 2\theta_n)\}_{n=1}^N$ vanishes. In fact, one can show that

$$\mathrm{FP}(F) - \tfrac{N^2}{2} = \Big(\sum_{n=1}^N \cos^2\theta_n\Big)^2 + 2\Big(\sum_{n=1}^N \cos\theta_n\sin\theta_n\Big)^2 + \Big(\sum_{n=1}^N \sin^2\theta_n\Big)^2 - \tfrac{N^2}{2} = \tfrac{1}{2}\Big[\Big(\sum_{n=1}^N \cos 2\theta_n\Big)^2 + \Big(\sum_{n=1}^N \sin 2\theta_n\Big)^2\Big],$$

and so $\|FF^* - \frac{N}{2}\mathrm{I}\|_{\mathrm{HS}} = \frac{1}{\sqrt{2}}\|\sum_{n=1}^N(\cos 2\theta, \sin 2\theta)\|$. That is, given any unit vectors in $\mathbb{R}^2$, double their polar angles, and add the resulting vectors, base-to-tip; for this chain of vectors, the distance between its head and tail is proportional to the original vectors' distance from tightness. In particular, our physical intuition tells us that if a collection of unit vectors is close to being tight, then their double-angle counterparts must only be slightly perturbed in order to close their chain, meaning the original vectors are indeed close to a UNTF. But how close? To begin to answer this question, consider the following example:

$$F(\theta) := \begin{bmatrix} \cos\theta & \cos\theta & 0 & 0 \\ \sin\theta & -\sin\theta & 1 & 1 \end{bmatrix}, \qquad \tilde{F}(\theta) := \begin{bmatrix} \cos\frac{\theta}{2} & \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} & \sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}. \tag{38}$$

One can show that $\|F(\theta)F^*(\theta) - \frac{N}{2}\mathrm{I}\|_{\mathrm{HS}}^2 = 8\sin^4\theta$, while $\sum_{n=1}^N \|P_n(\theta)F(\theta)F^*(\theta)f_n(\theta)\|^2 = 32\sin^6\theta\cos^2\theta$. That said, unlike in (18), there is no factor A independent of θ such that $A\|F(\theta)F^*(\theta) - \frac{N}{2}\mathrm{I}\|_{\mathrm{HS}}^2 \leq \sum_{n=1}^N \|P_n(\theta)F(\theta)F^*(\theta)f_n(\theta)\|^2$ for all θ. Therefore, at the very least, our analysis of the gradient descent algorithm, given in the previous section, must be refined in order to guarantee convergence.

Nevertheless, in this example, we can show that gradient descent does, in fact, converge to a UNTF, albeit at a sublinear rate. Here, $g_1(\theta) = 4\cos\theta\sin^3\theta(-\sin\theta, \cos\theta)$, $g_2(\theta) = -4\cos\theta\sin^3\theta(\sin\theta, \cos\theta)$, and $g_3(\theta) = g_4(\theta) = 0$. Recalling Proposition 1, one can show that $F(\theta; t) = F(\theta - 4t\cos\theta\sin^3\theta)$. That is, each iteration transforms an arrangement of angle θ into a new arrangement with angle $\theta - 4t\cos\theta\sin^3\theta$; repeated iterations indeed converge to $\theta = 0$, albeit very slowly. In this way, gradient descent converges to $\{e_1, e_1, e_2, e_2\}$, that is, two copies of the standard basis, which is indeed a UNTF. Note that since the limiting frame is OP, we know that for each $\varepsilon > 0$, the F_k's eventually become ε-OP—this is why the linear rate of convergence guaranteed by Theorem 6 does not hold here.

This same example can be used to give a baseline on answers to the Paulsen problem in the non-relatively-prime case. Indeed, noting that every real UNTF in $\mathbb{S}_2^4$ is the union of two orthonormal bases, we can show that for each $\theta \in [0, \frac{\pi}{8}]$, $\tilde{F}(\theta)$ is the closest UNTF to $F(\theta)$. But, $\|\tilde{F}(\theta) - F(\theta)\|_{\text{HS}} = 4 \sin \frac{\theta}{4}$, which is on the order of the square-root of $\|F(\theta)F^*(\theta) - \frac{N}{2}\text{I}\|_{\text{HS}}^{\frac{1}{2}}$ as θ grows small. As such, (38) is a counterexample to the sometimes-voiced belief that distance from a UNTF is at worst a linear function of distance from tightness. In other words, recalling (3), $\alpha = 1$ is not possible for every M and N; even when $M = 2$ and $N = 4$, the best possible α is $\frac{1}{2}$. This leads to three important questions: 1) For a given M and N, is the version of the Paulsen problem given in (3) even solvable? 2) If so, what is the best possible α for a given M and N? 3) Is there a single α that works for all M and N, or does performance truly depend on the number of common factors between M and N? Below, we outline an argument that answers the first question in the affirmative; the second and third questions remain open.

As the preceeding example illustrated, gradient descent is not guaranteed to converge in the non-relatively-prime case, since there is no ε for which iterations never become ε-OP. To resolve this issue, we introduce the concept of "jumping" to a nearby OP unit norm frame:

Theorem 10. *Let $\varepsilon \in (0, \frac{1}{2M}]$. Then, for every ε-orthogonally partitionable $F \in \mathbb{S}_M^N$, there exists an orthogonally partitionable $\tilde{F} \in \mathbb{S}_M^N$ such that $\|\tilde{F} - F\|_{\text{HS}} \leq (2N)^{\frac{1}{2}}(M\varepsilon)^{\frac{1}{3}}$.*

Proof. We first claim that for every unit vector $f \in \mathbb{H}_M$ and every nonzero projection operator P on $\mathbb{H}_M$, there exists a unit vector $g \in P(\mathbb{H}_M)$ such that $\|f - g\|^2 \leq 2\|(I - P)f\|^2$. If $Pf = 0$, we may take g to be any unit vector in $P(\mathbb{H}_M)$, since that would mean $\|f - g\|^2 = 2 = 2\|f\|^2 = 2\|(I - P)f\|^2$.Otherwise, we take $g = \frac{Pf}{\|Pf\|}$, since

$$\left\|f - \tfrac{Pf}{\|Pf\|}\right\|^2 = \left\|Pf + (I - P)f - \tfrac{Pf}{\|Pf\|}\right\|^2 = \left\|(1 - \tfrac{1}{\|Pf\|})Pf + (I - P)f\right\|^2,$$

and so the Pythagorean theorem gives

$$\left\|f - \tfrac{Pf}{\|Pf\|}\right\|^2 = (1 - \tfrac{1}{\|Pf\|})^2\|Pf\|^2 + \|(I - P)f\|^2 = 2(1 - \|Pf\|) \leq 2(1 - \|Pf\|^2) = 2\|(I - P)f\|^2. \tag{39}$$

For simplicity, we take $g := \frac{Pf}{\|Pf\|}$, understanding what this means when $Pf = 0$.

Since F is ε-OP, we have $\mathcal{I} \sqcup \mathcal{J} = \{1, \ldots, N\}$ such that $|\langle f_i, f_j\rangle| < \varepsilon$ whenever $i \in \mathcal{I}$ and $j \in \mathcal{J}$. Without loss of generality, we take $|\mathcal{I}| \geq |\mathcal{J}|$. Defining $F_{\mathcal{I}} := \{f_i\}_{i\in\mathcal{I}}$, the frame operator $F_{\mathcal{I}}F_{\mathcal{I}}^*$ has eigenvalues $\{\lambda_{\mathcal{I},m}\}_{m=1}^M$, arranged in decreasing order, and eigenvectors $\{e_{\mathcal{I},m}\}_{m=1}^M$. Take $\lambda' := \frac{2N}{3}(\frac{\varepsilon^2}{M})^{\frac{1}{3}}$, and define $M_{\mathcal{I}} := \#\{m : \lambda_{\mathcal{I},m} \geq \lambda'\}$. We know $M_{\mathcal{I}} \geq 1$, since otherwise

$$\tfrac{N}{2} \leq |\mathcal{I}| = \text{Tr}(F_{\mathcal{I}}^*F_{\mathcal{I}}) = \text{Tr}(F_{\mathcal{I}}F_{\mathcal{I}}^*) = \sum_{m=1}^{M} \lambda_{\mathcal{I},m} < M\lambda' = \tfrac{2N}{3}(M\varepsilon)^{\frac{2}{3}} \leq \tfrac{2^{\frac{1}{3}}N}{3} < \tfrac{N}{2}.$$

Therefore, $P := \sum_{m=1}^{M_{\mathcal{I}}} e_{\mathcal{I},m}e_{\mathcal{I},m}^*$ is a nonzero projection operator on $\mathbb{H}_M$. Moreover,

$$\sum_{i\in\mathcal{I}} \|(I - P)f_i\|^2 = \sum_{i\in\mathcal{I}} \sum_{m=M_{\mathcal{I}}+1}^{M} |\langle f_i, e_{\mathcal{I},m}\rangle|^2 = \sum_{m=M_{\mathcal{I}}+1}^{M} \|F_{\mathcal{I}}^*e_{\mathcal{I},m}\|^2 = \sum_{m=M_{\mathcal{I}}+1}^{M} \langle F_{\mathcal{I}}F_{\mathcal{I}}^*e_{\mathcal{I},m}, e_{\mathcal{I},m}\rangle = \sum_{m=M_{\mathcal{I}}+1}^{M} \lambda_{\mathcal{I},m} < M\lambda'. \tag{40}$$

Also, the fact that $e_{\mathcal{I},m}$ is an eigenvector of $F_{\mathcal{I}}F_{\mathcal{I}}^*$ with eigenvalue $\lambda_{\mathcal{I},m}$ gives

$$\sum_{j\in\mathcal{J}} \|Pf_j\|^2 = \sum_{j\in\mathcal{J}} \sum_{m=1}^{M_{\mathcal{I}}} |\langle f_j, e_{\mathcal{I},m}\rangle|^2 = \sum_{j\in\mathcal{J}} \sum_{m=1}^{M_{\mathcal{I}}} \left|\left\langle f_j, \tfrac{1}{\lambda_{\mathcal{I},m}} \sum_{i\in\mathcal{I}} \langle e_{\mathcal{I},m}, f_i\rangle f_i\right\rangle\right|^2 \leq \sum_{j\in\mathcal{J}} \sum_{m=1}^{M_{\mathcal{I}}} \tfrac{1}{\lambda_{\mathcal{I},m}^2}\left(\sum_{i\in\mathcal{I}} |\langle e_{\mathcal{I},m}, f_i\rangle||\langle f_i, f_j\rangle|\right)^2.$$

Continuing, we use $|\langle f_i, f_j\rangle| \leq \varepsilon$ and $\lambda_{\mathcal{I},m} \geq \lambda'$:

$$\sum_{j\in\mathcal{J}} \|Pf_j\|^2 \leq \tfrac{\varepsilon^2}{(\lambda')^2} \sum_{j\in\mathcal{J}} \sum_{m=1}^{M_{\mathcal{I}}}\left(\sum_{i\in\mathcal{I}} |\langle e_{\mathcal{I},m}, f_i\rangle|\right)^2 \leq \tfrac{\varepsilon^2}{(\lambda')^2}|\mathcal{I}| \sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{J}} \sum_{m=1}^{M_{\mathcal{I}}} |\langle e_{\mathcal{I},m}, f_i\rangle|^2 \leq \tfrac{\varepsilon^2}{(\lambda')^2}|\mathcal{I}|^2|\mathcal{J}| \leq \tfrac{4N^3\varepsilon^2}{27(\lambda')^2}, \tag{41}$$

where the last inequality comes from $|\mathcal{I}|^2(N-|\mathcal{I}|) \leq \frac{4N^3}{27}$. Define $\tilde{F} = \{\tilde{f}_n\}_{n=1}^N$ by $\tilde{f}_n = \frac{Pf_n}{\|Pf_n\|}$ when $n \in \mathcal{I}$, and $\tilde{f}_n = \frac{(I-P)f_n}{\|(I-P)f_n\|}$ when $n \in \mathcal{J}$. Then, combining (39) with (40) and (41) gives the result:

$$\|\tilde{F}-F\|_{\text{HS}}^2 = \sum_{i\in\mathcal{I}}\left\|f_i - \tfrac{Pf_i}{\|Pf_i\|}\right\|^2 + \sum_{j\in\mathcal{J}}\left\|f_j - \tfrac{(I-P)f_j}{\|(I-P)f_j\|}\right\|^2 \leq \sum_{i\in\mathcal{I}} 2\|(I-P)f_i\|^2 + \sum_{j\in\mathcal{J}} 2\|Pf_j\|^2 < 2M\lambda' + \tfrac{8N^3\varepsilon^2}{27(\lambda')^2} = 2N(M\varepsilon)^{\frac{2}{3}}. \quad \square$$

The previous result tells us how far we must jump in order to transform an ε-OP frame into one that is exactly OP. This opens the door for the following procedure for producing UNTFs in the non-relatively-prime case: given a collection of unit norm vectors and fixing any $\varepsilon \in (0,1]$, perform gradient descent until one's vectors become ε-OP, at which jump to a OP frame, and then repeat this procedure on each of the two subframes. In the following result, we use Theorems 6 and 10 to bound how far this procedure will take us from our original frame.

Theorem 11. *Suppose M and N are not relatively prime. Take $F \in \mathbb{S}_M^N$ such that $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\text{HS}} \leq (2^{21}M^{27}N^{14})^{-1}$. Then there exists $\tilde{F} \in \mathbb{S}_M^N$, which is either a unit norm tight frame or is orthogonally partitionable, with equal redundancies in each of the two partitioned subspaces, such that*

$$\|\tilde{F}-F\|_{\text{HS}} \leq 3M^{\frac{6}{7}}N^{\frac{1}{2}}\left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\text{HS}}^{\frac{1}{7}}. \tag{42}$$

Proof. Take $t := \frac{1}{4N}$ and $\varepsilon := 2^{\frac{3}{2}}3^{\frac{3}{7}}M^{\frac{11}{7}}\|FF^* - \frac{M}{N}\mathrm{I}\|_{\text{HS}}^{\frac{3}{7}}$. According to Theorem 6, gradient descent will converge to a UNTF, provided iterations never become ε-OP. In this way, we either converge to a UNTF $\tilde{F}$, or produce an ε-OP frame within $(2N)^{\frac{1}{2}}(M\varepsilon)^{\frac{1}{3}}$ of an OP frame $\tilde{F}$, by Theorem 10.Either way, Theorems 6 and 10 give

$$\|\tilde{F}-F\|_{\text{HS}} \leq \tfrac{8M^4N^{\frac{1}{2}}}{\varepsilon^2}\left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\text{HS}} + (2N)^{\frac{1}{2}}(M\varepsilon)^{\frac{1}{3}} = 3^{-\frac{6}{7}}7M^{\frac{6}{7}}N^{\frac{1}{2}}\left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\text{HS}}^{\frac{1}{7}},$$

which proves (42). Now suppose $\tilde{F}$ is OP. Since

$$\begin{aligned}\left|\text{FP}(\tilde{F}) - \text{FP}(F)\right| &= \text{Tr}[(\tilde{F}\tilde{F}^* - FF^*)(\tilde{F}\tilde{F}^* + FF^*)] \\ &\leq \|\tilde{F}\tilde{F}^* - FF^*\|_{\text{HS}}\|\tilde{F}\tilde{F}^* + FF^*\|_{\text{HS}} \\ &\leq \|\tilde{F}-F\|_{\text{HS}}\big(\|\tilde{F}\|_{\text{HS}} + \|F\|_{\text{HS}}\big)\big(\|\tilde{F}\|_{\text{HS}}^2 + \|F\|_{\text{HS}}^2\big),\end{aligned}$$

we use $\|F\|_{\text{HS}}^2 = \|\tilde{F}\|_{\text{HS}}^2 = N$ to get $|\text{FP}(\tilde{F}) - \text{FP}(F)| \leq 4N^{\frac{3}{2}}\|\tilde{F}-F\|_{\text{HS}}$. Therefore,

$$\text{FP}(\tilde{F}) \leq \text{FP}(F) + \left|\text{FP}(\tilde{F}) - \text{FP}(F)\right| = \tfrac{N^2}{M} + \left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\text{HS}}^2 + \left|\text{FP}(\tilde{F}) - \text{FP}(F)\right| \leq \tfrac{N^2}{M} + \left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\text{HS}}^2 + 4N^{\frac{3}{2}}\|\tilde{F}-F\|_{\text{HS}}.$$

Continuing, we apply (42) and use the fact that $\|FF^* - \frac{N}{M}\mathrm{I}\|_{\text{HS}}^2 \leq 4N^{\frac{3}{2}}(3M^{\frac{6}{7}}N^{\frac{1}{2}}\|FF^* - \frac{N}{M}\mathrm{I}\|_{\text{HS}}^{\frac{1}{7}})$:

$$\text{FP}(\tilde{F}) \leq \tfrac{N^2}{M} + \left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\text{HS}}^2 + 4N^{\frac{3}{2}}\left(3M^{\frac{6}{7}}N^{\frac{1}{2}}\left\|FF^* - \tfrac{N}{M}\mathrm{I}\right\|_{\text{HS}}^{\frac{1}{7}}\right) \leq \tfrac{N^2}{M} + \tfrac{24M^{\frac{6}{7}}N^2}{(2^{21}M^{27}N^{14})^{\frac{1}{7}}} = \tfrac{N^2}{M} + \tfrac{3}{M^3}. \tag{43}$$

Since $\tilde{F}$ is OP, there exists an orthogonal partition $\mathcal{I} \sqcup \mathcal{J} = \{1,\dots,N\}$. Take $M_{\mathcal{I}}$ to be the dimension of the span of $\{f_n\}_{n\in\mathcal{I}}$. Then,

$$\text{FP}(\tilde{F}) = \text{FP}(\tilde{F}_{\mathcal{I}}) + \text{FP}(\tilde{F}_{\mathcal{J}}) \geq \tfrac{|\mathcal{I}|^2}{M_{\mathcal{I}}} + \tfrac{(N-|\mathcal{I}|)^2}{M-M_{\mathcal{I}}} = \tfrac{N^2}{M} + \tfrac{(|\mathcal{I}|M-M_{\mathcal{I}}N)^2}{MM_{\mathcal{I}}(M-M_{\mathcal{I}})}.$$

In particular, if $|\mathcal{I}|M - M_{\mathcal{I}}N \neq 0$, then $(|\mathcal{I}|M - M_{\mathcal{I}}N)^2 \geq 1$, and since $M_{\mathcal{I}}(M-M_{\mathcal{I}}) \leq \frac{M}{4}$, we would have $\text{FP}(\tilde{F}) \geq \frac{N^2}{M} + \frac{4}{M^3}$. Considering (43), we may conclude that $|\mathcal{I}|M - M_{\mathcal{I}}N = 0$, and so $\frac{N}{M} = \frac{|\mathcal{I}|}{M_{\mathcal{I}}} = \frac{N-|\mathcal{I}|}{M-M_{\mathcal{I}}}$. $\square$

Repeated applications of Theorem 11 will provide solutions, albeit inelegant ones, to the Paulsen problem given in (3). To elaborate, Theorem 11 states that if a unit norm frame F is sufficiently tight, then there exists a unit norm $\tilde{F}$ such that $\|\tilde{F}-F\|_{\text{HS}} = \text{O}(\|FF^* - \frac{N}{M}\mathrm{I}\|^{\frac{1}{7}})$ which is either a UNTF or is OP into components of equal redundancy. Since we are done if $\tilde{F}$ happens to be a UNTF, let's focus on the case where $\tilde{F}$ is OP, that is, when $\tilde{F} = \tilde{F}_{\mathcal{I}} \oplus \tilde{F}_{\mathcal{J}}$, where $\tilde{F}_{\mathcal{I}} = \{\tilde{f}_i\}_{i\in\mathcal{I}}$ and $\tilde{F}_{\mathcal{J}} = \{\tilde{f}_j\}_{j\in\mathcal{J}}$ are frames for some $M_{\mathcal{I}}$- and $M_{\mathcal{J}}$-dimensional subspaces of $\mathbb{H}_M$, respectively, and

$\frac{|\mathcal{I}|}{M_\mathcal{I}} = \frac{|\mathcal{J}|}{M_\mathcal{J}} = \frac{N}{M}$. We then apply Theorem 11 to $\tilde{F}_\mathcal{I}$ and $\tilde{F}_\mathcal{J}$: if each is close to a UNTF, these can be directly summed to form a UNTF which is close to $\tilde{F}$ and in turn, to F; if either is OP, we must continue this process in lower-dimensional subspaces. At most M such nested applications of Theorem 11 are necessary, since each reduces the dimension of the space in consideration by at least 1. The main issue is that each application of Theorem 11 comes at a terrible cost: "jumping" from an ε-OP sequence to an OP sequence can increase one's frame potential by a constant multiple of the jump distance. In particular, with each application of Theorem 11, one's distance from tightness may be effectively raised to a $\frac{1}{7}$ power; when one's distance is very small, this exponentiation results in a dramatic increase in distance. When applied M times in succession, one would therefore expect a net exponent of $\frac{1}{7^M}$. That is, we expect that there exists an extremely small $\delta > 0$ and an extremely large C for which (3) will hold for $\alpha = \frac{1}{7^M}$. It is unknown whether such an M-dependent α is inherent to this problem, or simply a consequence of a weak argument on our part.

We emphasize that such issues, while of great mathematical interest, should cause little worry in real-world applications. Indeed, the "perform gradient descent and jump when approaching OP" method that we employed in the proof of Theorem 11 produces UNTFs which, for all practical purposes, are close to their originals. Nevertheless, the issue stands: this distance may not be a nice function of the tightness itself. Indeed, this is the heart of the part of the Paulsen problem that remains open: "Given a unit norm frame which is extremely close to being tight, and is also extremely close to being OP, how far away, as a function of tightness, is the nearest UNTF?" This problem reveals our current lack of understanding of the geometry of the set of all UNTFs on very small neighborhoods of OP UNTFs, and is more than worthy of additional study.

Acknowledgments

Casazza was supported by NSF DMS 0704216 and 1008183. Fickus was supported by AFOSR F1ATA09125G003. The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

References

[1] J. J. Benedetto, M. Fickus, Finite normalized tight frames, Adv. Comput. Math. 18 (2003) 357–385.
[2] B. G. Bodmann, P. G. Casazza, The road to equal-norm Parseval frames, J. Funct. Anal. 258 (2010) 397–420.
[3] P. G. Casazza, M. Fickus, Minimizing fusion frame potential, Acta Appl. Math. 107 (2009) 7–24.
[4] P. G. Casazza, M. Fickus, Gradient descent of the frame potential, Proc. Sampl. Theory Appl. (2009), 1–4.
[5] P. G. Casazza, M. Fickus, J. Kovačević, M.T. Leon, J. C. Tremain, A physical interpretation of tight frames, in: Harmonic Analysis and Applications: In Honor of John J. Benedetto, C. Heil ed., Birkhäuser, Boston, pp. 51–76 (2006).
[6] P. G. Casazza, M. Fickus, D. G. Mixon, Y. Wang, Z. Zhou, Constructing tight fusion frames, to appear in: Appl. Comput. Harmon. Anal.
[7] P. G. Casazza, J. Kovačević, Equal-norm tight frames with erasures, Adv. Comp. Math. 18 (2003) 387–430.
[8] P. G. Casazza, M. Leon, Existence and construction of finite tight frames, J. Comput. Appl. Math. 4 (2006) 277–289.
[9] A. Chebira, M. Fickus, D. G. Mixon, Filter bank fusion frames, submitted.
[10] K. Dykema, N. Strawn, Manifold structure of spaces of spherical tight frames, Int. J. Pure Appl. Math. 28 (2006) 217–256.
[11] M. Fickus, B. D. Johnson, K. Kornelson, K. Okoudjou, Convolutional frames and the frame potential, Appl. Comput. Harmon. Anal. 19 (2005) 77–91.
[12] V. K. Goyal, J. Kovačević, J. A. Kelner, Quantized frame expansions with erasures, Appl. Comput. Harmon. Anal. 10 (2001) 203–233.
[13] V. K. Goyal, M. Vetterli, N. T. Thao, Quantized overcomplete expansions in $\mathbb{R}^N$: Analysis, synthesis, and algorithms, IEEE Trans. Inform. Theory 44 (1998) 16–31.
[14] R. B. Holmes, V. I. Paulsen, Optimal frames for erasures, Linear Algebra Appl. 377 (2004) 31–51.
[15] B. D. Johnson, K. Okoudjou, Frame potential and finite abelian groups, Contemp. Math. 464 (2008) 137–148.
[16] J. Kovačević, A. Chebira, Life beyond bases: The advent of frames (Part I), IEEE Signal Process. Mag. 24 (2007) 86–104.
[17] J. Kovačević, A. Chebira, Life beyond bases: The advent of frames (Part II), IEEE Signal Process. Mag. 24 (2007) 115–125.
[18] P. Massey, Optimal reconstruction systems for erasures and for the q-potential, Linear Algebra Appl. 431 (2009) 1302–1316.
[19] P. Massey, M. Ruiz, Minimization of convex functionals over frame operators, Adv. Comput. Math. 32 (2010) 131–153.
[20] P. Massey, M. Ruiz, D. Stojanoff, The structure of minimizers of the frame potential on fusion frames, to appear in J. Fourier Anal. Appl.