

A Gramian Description of the Degree 4 Generalized Elliptope

Afonso S. Bandeira*

Dmitriy Kunisky†

First Draft: December 30, 2018

Current Draft: March 24, 2019

Abstract

One of the most widely studied convex relaxations in combinatorial optimization is the relaxation of the cut polytope \mathcal{C}^N to the elliptope \mathcal{E}^N , which corresponds to the degree 2 sum-of-squares (SOS) relaxation of optimizing a quadratic form over the hypercube $\{\pm 1\}^N$. We study the extension of this classical idea to degree 4 SOS, which gives an intermediate relaxation we call the *degree 4 generalized elliptope* \mathcal{E}_4^N . Our main result is a necessary and sufficient condition for the Gram matrix of a collection of vectors to belong to \mathcal{E}_4^N . Consequences include a tight rank inequality between degree 2 and degree 4 pseudomoment matrices, and a guarantee that the only extreme points of \mathcal{E}^N also in \mathcal{E}_4^N are the cut matrices; that is, \mathcal{E}^N and \mathcal{E}_4^N share no “spurious” extreme point.

For Gram matrices of equiangular tight frames, we give a simple criterion for membership in \mathcal{E}_4^N . This yields new inequalities satisfied in \mathcal{E}_4^N but not \mathcal{E}^N whose structure is related to the Schläfli graph and which cannot be obtained as linear combinations of triangle inequalities. We also give a new proof of the restriction to degree 4 of a result of Laurent showing that \mathcal{E}_4^N does not satisfy certain cut polytope inequalities capturing parity constraints. Though limited to this special case, our proof of the positive semidefiniteness of Laurent’s pseudomoment matrix is short and elementary.

Our techniques also suggest that membership in \mathcal{E}_4^N is closely related to the partial transpose operation on block matrices, which has previously played an important role in the study of quantum entanglement. To illustrate, we present a correspondence between certain entangled bipartite quantum states and the matrices of $\mathcal{E}_4^N \setminus \mathcal{C}^N$.

*Courant Institute of Mathematical Sciences and Center for Data Science, New York University, NY 10012. Afonso S. Bandeira was partially supported by NSF grants DMS-1712730 and DMS-1719545, and by a grant from the Sloan Foundation.

†Courant Institute of Mathematical Sciences, New York University, NY 10012. Dmitriy Kunisky was partially supported by NSF grants DMS-1712730 and DMS-1719545.

Contents

1	Introduction	3
2	Main Results	5
2.1	Preliminaries	5
2.2	Gramian Description of \mathcal{E}_4^N	8
2.3	Constraints on Pseudomoment Extensions	9
2.4	Examples from Equiangular Tight Frames	10
2.5	Applications	11
3	Notations	13
4	Gramian Description of \mathcal{E}_4^N: Theorem 2.15	14
4.1	Proof of Theorem 2.15: Pseudomoment Witness to Gram Vector Witness	14
4.2	Proof of Theorem 2.15: Gram Vector Witness to Pseudomoment Witness	16
4.3	Interpreting Theorem 2.15 as a Relaxation	18
5	Constraints on Witnesses: Lemma 2.16 and Theorem 2.17	19
5.1	Proof of Lemma 2.16	19
5.2	Proof of Theorem 2.17	22
6	Examples from Equiangular Tight Frames: Theorem 2.19	23
6.1	Proof of Theorem 2.19	23
7	Applications	25
7.1	Schläfli Inequalities: Theorem 2.20	25
7.2	Complexity of Parity Inequalities	26
A	Pseudomoment Reductions for Sum-of-Squares over $\{\pm 1\}^N$	32
B	Proofs of Structural Results on $\mathcal{B}(N, r)$	33
B.1	Proof of Proposition 4.1	33
B.2	Proof of Proposition 4.2	34
C	Proofs of Relaxation Descriptions of Theorem 2.15	35
C.1	Proof of Proposition 4.3	35
C.2	Proof of Proposition 4.5	36
D	Proofs of Results on Partial Transposition	38
D.1	Proof of Proposition 5.1	38
D.2	Proof of Proposition 5.2	39
D.3	Proof of Proposition 5.3	39
E	Tight Frame Projector Calculations	41
E.1	Projector to V_{sym}	42
E.2	Projector to V'_{sym}	42

1 Introduction

The optimization of quadratic forms over the hypercube $\{\pm 1\}^N$,

$$M(\mathbf{W}) := \max_{\mathbf{x} \in \{\pm 1\}^N} \mathbf{x}^\top \mathbf{W} \mathbf{x}, \quad (1)$$

is a well-studied computational problem arising in several contexts, including the Grothendieck problem [33, 42], graph problems such as finding the maximum cut [29, 62], synchronization over the group $\mathbb{Z}/2\mathbb{Z}$ [1, 4], spiked matrix estimation problems under priors with the i.i.d. Rademacher distribution [2, 61], and the determination of ground state energies of hard two-spin models from statistical physics [5, 58]. $M(\mathbf{W})$ may equivalently be viewed as optimizing a linear objective over a convex set called the *cut polytope*:

$$\mathcal{C}^N := \text{conv}(\{\mathbf{x}\mathbf{x}^\top : \mathbf{x} \in \{\pm 1\}^N\}), \quad (2)$$

$$M(\mathbf{W}) = \max_{\mathbf{X} \in \mathcal{C}^N} \langle \mathbf{W}, \mathbf{X} \rangle. \quad (3)$$

Though it is convex, this problem is nonetheless difficult to solve exactly (e.g., NP-hard for \mathbf{W} a graph Laplacian, which computes the maximum cut [40]) due to the intricate discrete geometry of the cut polytope [22].

A popular algorithmic choice for approximating $M(\mathbf{W})$ and estimating its optimizer is to form *relaxations* of \mathcal{C}^N , larger convex sets admitting simpler descriptions. Often, the relaxed sets may be described concisely in terms of positive semidefiniteness (psd) conditions, which leads to semidefinite programming (SDP) relaxations of $M(\mathbf{W})$. The most common way to execute this strategy is to optimize over the *elliptope*,

$$\mathcal{E}^N = \mathcal{E}_2^N := \{\mathbf{X} \in \mathbb{R}_{\text{sym}}^{N \times N} : \mathbf{X} \succeq 0, \text{diag}(\mathbf{X}) = \mathbf{1}\} \supseteq \mathcal{C}^N. \quad (4)$$

For example, the well-known approximation algorithms of Goemans-Williamson [29] and Nesterov [57] optimize over \mathcal{E}_2^N and then perform a *rounding* procedure to recover an approximately optimal $\mathbf{x} \in \{\pm 1\}^N$ from $\mathbf{X} \in \mathcal{E}_2^N$.

As our notation suggests, \mathcal{E}_2^N is only the first of a sequence of increasingly tighter relaxations of \mathcal{C}^N , corresponding to *sum-of-squares (SOS)* relaxations of $M(\mathbf{W})$. These sets are indexed by an even integer $d \geq 2$ called the *degree*, and we call the set described at degree d the *degree d generalized elliptope*, denoted \mathcal{E}_d^N . The generalized elliptopes satisfy the inclusions

$$\mathcal{E}_2^N \supseteq \mathcal{E}_4^N \supseteq \cdots \supseteq \mathcal{E}_{N+1\{N \text{ odd}\}}^N = \mathcal{C}^N. \quad (5)$$

(The last equality and its tightness in the sense that no generalized elliptope of lower degree equals \mathcal{C}^N are non-trivial results proven by [24] and [46] respectively.) Thus, optimizing over generalized elliptopes of higher degree may yield better approximations of $M(\mathbf{W})$; however, it is also costlier, since the associated semidefinite program is over matrix variables of size $N^{d/2} \times N^{d/2}$, whereby while general-purpose SDP algorithms will solve such an optimization to fixed accuracy in polynomial time in N , the bound on their runtime will be of order $N^{O(d)}$ [8]. It is therefore important to know whether optimizing over generalized elliptopes of constant degree $d > 2$ actually improves the bounds on $M(\mathbf{W})$ achieved by optimizing over \mathcal{E}_2^N on specific classes of optimization problems as $N \rightarrow \infty$.¹

¹There is an extensive literature relating this question to the Unique Games Conjecture [43, 69], which implies for several problems, most notably MaxCut, that optimizing over generalized elliptopes of constant degree cannot improve the worst-case approximation ratio achieved by optimizing over \mathcal{E}_2^N (see e.g. [41, 63]). On the other hand, similar questions in the average case, for instance the typical quality of approximation that can be achieved for natural random models of \mathbf{W} , are only beginning to be understood. For instance, [56] is a recent work in this direction concerning SDP relaxations for MaxCut and related problems on random graphs.

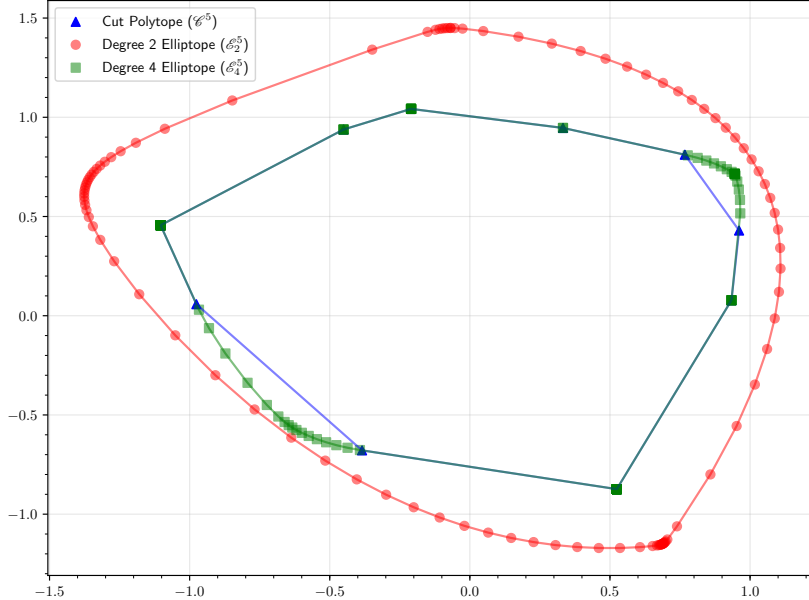


Figure 1: **Cut polytope and elliptopes in low dimension.** We plot the cross-section of \mathcal{C}^5 , \mathcal{E}_2^5 , and \mathcal{E}_4^5 by an isotropic random subspace (in the off-diagonal matrix entries) by numerically solving suitably constrained linear and semidefinite programs. This is different from the projection of these sets onto such a subspace, for which we observe that \mathcal{E}_4^5 is almost always indistinguishable from \mathcal{C}^5 . Note also that $N = 5$ is the lowest dimension where $\mathcal{E}_4^N \neq \mathcal{C}^N$.

Nonetheless, although many fundamental geometric results on \mathcal{E}_2^N were obtained soon after this relaxation was introduced [48, 49], relatively little remains known about the geometry of \mathcal{E}_d^N for specific constant values of $d > 2$. Instead, most progress on higher-degree SOS relaxations has been through the algebraic interpretation of SOS, and concerning the limit $d \rightarrow \infty$ after $N \rightarrow \infty$. A prominent line of work in this direction is on negative results for SOS, proving that its approximations are poor for some problems even at high degrees. This began with the works of Grigoriev [30, 31, 32] giving lower bounds for the SOS degree needed prove linear systems over $\mathbb{Z}/2\mathbb{Z}$ unsatisfiable. With similar techniques, Laurent [46] produced examples of inequalities over \mathcal{C}^N that do not hold over \mathcal{E}_d^N until $d = \Omega(N)$. Schoenebeck [65] later rediscovered Grigoriev’s results and emphasized their application to constraint satisfaction problems. More recently, similar ideas yielded rapid progress on the planted clique problem [55, 21, 35, 6], culminating in a framework called *pseudo-calibration* introduced in [6] for constructing pseudomoment matrices based on statistical reasoning, which has since been applied to several other problems [36, 64]. The general proof technique behind these results is to first construct candidate pseudomoment matrices entrywise, and then analyze whether those candidates satisfy the necessary constraints. Among those constraints, the requirement that the pseudomoment matrix be psd is notoriously difficult to verify.

In this paper, we make two contributions to the state of affairs outlined above. First, we derive some novel geometric facts about \mathcal{E}_4^N , the first generalized elliptope, relating its extrema to the facial geometry of \mathcal{E}_2^N . Second, in doing so, we introduce a technique for constructing SOS pseudomoment matrices as Gram matrices of collections of vectors, which are therefore guaranteed by construction to be psd. We use this to provide an alternate proof of part of Laurent’s theorem [46] mentioned above, and believe that this idea may eventually lead to simplified proofs of other difficult negative results in the SOS optimization literature.

2 Main Results

2.1 Preliminaries

Pseudomoment matrices. We first present the description of \mathcal{E}_d^N in terms of the *pseudomoment* interpretation of SOS optimization. The formal definition is as follows.

Definition 2.1. For a finite set \mathcal{A} , we write $\mathcal{A}^{\leq d}$ for the collection of finite (possibly empty) strings in the elements of \mathcal{A} of length at most d , i.e.

$$\mathcal{A}^{\leq d} := \{\emptyset\} \sqcup \mathcal{A} \sqcup \mathcal{A}^2 \sqcup \cdots \sqcup \mathcal{A}^d. \quad (6)$$

The sets $\mathcal{A}^{\leq d}$ for all $d \in \mathbb{N} := \{n \in \mathbb{Z} : n \geq 0\}$ may be thought of as embedded in the set $\mathcal{A}^{< \infty}$ of all finite strings in the elements of \mathcal{A} . For $\mathbf{s} \in \mathcal{A}^{< \infty}$, we write $|\mathbf{s}|$ for the length of \mathbf{s} , and for $\mathbf{s}, \mathbf{t} \in \mathcal{A}^{< \infty}$, we write $\mathbf{s} \circ \mathbf{t} \in \mathcal{A}^{|\mathbf{s}|+|\mathbf{t}|}$ for the concatenation of \mathbf{s} and \mathbf{t} .

Definition 2.2. For $\mathbf{s} \in \mathcal{A}^{< \infty}$, $\text{odd}(\mathbf{s}) \subset \mathcal{A}$ denotes the set of symbols that occur an odd number of times in \mathbf{s} .

Definition 2.3. $\mathcal{E}_d^N \subset \mathbb{R}_{\text{sym}}^{N \times N}$ is the set of \mathbf{X} such that there exists $\mathbf{Y} \in \mathbb{R}^{N^{d/2} \times N^{d/2}}$, whose row and column indices we identify with the set $[N]^{d/2}$ ordered lexicographically, having $Y_{(1 \dots 1i)(1 \dots 1j)} = X_{ij}$ for all $i, j \in [N]$ and satisfying the following properties:

1. $\mathbf{Y} \succeq \mathbf{0}$.
2. $Y_{\mathbf{st}}$ only depends on $\text{odd}(\mathbf{s} \circ \mathbf{t})$.
3. $Y_{\mathbf{st}} = 1$ whenever $\text{odd}(\mathbf{s} \circ \mathbf{t}) = \emptyset$.

In this case, we say \mathbf{Y} is a degree d pseudomoment matrix over $\{\pm 1\}^N$ (or, more properly, for the constraints $\{x_i^2 - 1 = 0 : i \in [N]\}$) which extends \mathbf{X} .

(In Appendix A we present the standard reductions that allow us to restrict our attention to only the pseudomoments of degree exactly d for the case of relaxing the problem $\mathbf{M}(\mathbf{W})$.) In this paper, for the sake of brevity, we will simply refer to such \mathbf{Y} as a *degree d pseudomoment matrix*, since we only study optimization over $\{\pm 1\}^N$.

We will also only study degree 4 pseudomoment matrices in detail (though it will occasionally be instructive to refer to higher degree cases for comparison), so we briefly give a more concrete version of the above conditions for that case.

Proposition 2.4. Let $\mathbf{Y} \in \mathbb{R}^{N^2 \times N^2}$, with the row and column indices of \mathbf{Y} identified with pairs $(ij) \in [N]^2$ ordered lexicographically. Then, \mathbf{Y} is a degree 4 pseudomoment matrix if and only if the following conditions hold:

1. $\mathbf{Y} \succeq \mathbf{0}$.
2. $Y_{(ij)(kk)}$ does not depend on the index k .
3. $Y_{(ii)(ii)} = 1$ for every $i \in [N]$.
4. $Y_{(ij)(k\ell)}$ is invariant under permutations of the indices i, j, k, ℓ .

The intuitive meaning of these conditions is that \mathbf{Y} contains *pseudomoments* of degree 4 of a fictitious distribution over $\mathbf{x} \in \{\pm 1\}^N$, entry $Y_{(ij)(k\ell)}$ equaling the *pseudoexpectation* of $x_i x_j x_k x_\ell$. We represent the constraint $\mathbf{x} \in \{\pm 1\}^N$ as the system of polynomial constraints $x_i^2 = 1$ for $i \in [N]$, which constrains the pseudomoments per Conditions 2 and 3 of the definition. The remaining constraints are properties that the moments of probability distributions in general must satisfy. Conditions 1 through 4 taken together, however, still do not imply that a distribution over $\{\pm 1\}^N$ exists whose moments equal the specified pseudomoments.

The matrix \mathbf{X} , a minor of \mathbf{Y} , contains the degree 2 pseudomoments (which, as also discussed in greater detail in Appendix A, are in this case already among the degree 4 pseudomoments). We may then think of \mathcal{E}_4^N relaxing \mathcal{C}^N in the sense that \mathcal{C}^N is the set of degree 2 moment matrices of true distributions over $\{\pm 1\}^N$, while \mathcal{E}_4^N is the set of degree 2 pseudomoment matrices that “admit a consistent extension to degree 4.”

Note that the case $d > 2$ is fundamentally different from $d = 2$ in the important regard that, while \mathcal{E}_2^N is itself an affine slice of the psd cone (making it a so-called *spectrahedron*), there does not appear to be a straightforward way to describe any \mathcal{E}_d^N with $d > 2$ in this way. (On the other hand, it does not appear to be established that \mathcal{E}_d^N actually fails to be a spectrahedron for $d > 2$, though this is a natural conjecture. Some similar results in simpler cases are presented in [11].) Rather, these sets are most directly described as projections of spectrahedra (so-called *spectrahedral shadows*), the sets of all degree d pseudomoment matrices, from a higher-dimensional space. Thus, we should expect the generalized elliptopes to have a more subtle geometry than the classical elliptope, and our subject \mathcal{E}_4^N is the first of these subtler cases.

No more than the above is needed to understand our results, but for a more general and detailed presentation of the pseudomoment-and-pseudoexpectation optimization framework we direct the reader to [7, 45, 47].

Convex geometry. We next recall some basic notions from the geometry of convex sets. In what follows, let $K \subseteq \mathbb{R}^d$ be a closed convex set.

Definition 2.5. *The dimension of K is the dimension of the affine hull of K , denoted $\dim(K)$.*

Definition 2.6. *A convex subset $F \subseteq K$ is a face of K if whenever $\theta \mathbf{X} + (1 - \theta) \mathbf{Y} \in F$ with $\theta \in (0, 1)$ and $\mathbf{X}, \mathbf{Y} \in K$, then $\mathbf{X}, \mathbf{Y} \in F$.*

Definition 2.7. *$\mathbf{X} \in K$ is an extreme point of K if $\{\mathbf{X}\}$ is a face of K (of dimension zero).*

Definition 2.8. *The intersection of all faces of K containing $\mathbf{X} \in K$ is the unique smallest face of K containing \mathbf{X} , denoted $\text{face}_K(\mathbf{X})$.*

Definition 2.9. *The perturbation of \mathbf{X} in K is the subspace*

$$\text{pert}_K(\mathbf{X}) := \left\{ \mathbf{A} \in \mathbb{R}^d : \mathbf{X} \pm t\mathbf{A} \in K \text{ for all } t > 0 \text{ sufficiently small} \right\}. \quad (7)$$

The perturbation will come up naturally in our results, so we present the following useful fact giving its connection to the more intuitive objects from facial geometry.

Proposition 2.10. *Let $\mathbf{X} \in K$. Then,*

$$\text{face}_K(\mathbf{X}) = K \cap (\mathbf{X} + \text{pert}_K(\mathbf{X})). \quad (8)$$

In particular, the affine hull of $\text{face}_K(\mathbf{X})$ is $\mathbf{X} + \text{pert}_K(\mathbf{X})$, and therefore

$$\dim(\text{face}_K(\mathbf{X})) = \dim(\text{pert}_K(\mathbf{X})) \quad (9)$$

(in which there is a harmless reuse of notation between the dimension of a convex set and the dimension of a subspace).

Proof. Let $G := K \cap (\mathbf{X} + \text{pert}_K(\mathbf{X}))$. It is simple to check that $\mathbf{Y} \in G$ if and only if there exists $\mathbf{Y}' \in K$ and $\theta \in (0, 1]$ with $\mathbf{X} = \theta\mathbf{Y} + (1 - \theta)\mathbf{Y}'$ (and if $\theta < 1$ then $\mathbf{Y}' \in G$ as well).

Then, if F is any face of K containing \mathbf{X} , and $\mathbf{Y} \in G$, there exists $\mathbf{Y}' \in K$ and $\theta \in (0, 1]$ such that $\mathbf{X} = \theta\mathbf{Y} + (1 - \theta)\mathbf{Y}'$. If $\theta = 1$, then $\mathbf{Y} = \mathbf{X} \in F$. Otherwise, $\mathbf{Y} \in F$ by the definition of a face. Thus, in any case $\mathbf{Y} \in F$, so $G \subseteq F$. Since this holds for any face F containing \mathbf{X} , in fact $G \subseteq \text{face}_K(\mathbf{X})$.

It then suffices to show that G is a face of K . Suppose $\mathbf{Y} \in G$, and $\mathbf{Y}_1, \mathbf{Y}_2 \in K$ and $\theta \in (0, 1)$ with $\mathbf{Y} = \theta\mathbf{Y}_1 + (1 - \theta)\mathbf{Y}_2$. Since $\mathbf{Y} \in G$, there exists $\mathbf{Z} \in K$ and $\phi \in (0, 1]$ such that

$$\begin{aligned} \mathbf{X} &= \phi\mathbf{Y} + (1 - \phi)\mathbf{Z} \\ &= \phi(\theta\mathbf{Y}_1 + (1 - \theta)\mathbf{Y}_2) + (1 - \phi)\mathbf{Z} \\ &= \phi\theta\mathbf{Y}_1 + \phi(1 - \theta)\mathbf{Y}_2 + (1 - \phi)\mathbf{Z}. \end{aligned} \quad (10)$$

This is a convex combination of three points where the coefficients of \mathbf{Y}_1 and \mathbf{Y}_2 are strictly positive, so by the previous characterization we have $\mathbf{Y}_1, \mathbf{Y}_2 \in G$, completing the proof. \square

Finite-dimensional frames. Finally, we review some definitions of special types of *frames* in finite dimension, which are overcomplete collections of vectors with certain favorable geometric properties. A more thorough introduction, in particular for the more typical applications of these definitions in signal processing and harmonic analysis, may be found in [15]. In what follows, let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$ be unit vectors and let $\mathbf{X} := \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$ (meaning that $X_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ for $i, j \in [N]$).

Definition 2.11. *The vectors \mathbf{v}_i form a unit norm tight frame (UNTF) if any of the following equivalent conditions hold:*

1. $\sum_{i=1}^N \mathbf{v}_i \mathbf{v}_i^\top = \frac{N}{r} \mathbf{I}_r$.
2. *The eigenvalues of \mathbf{X} all equal either zero or $\frac{N}{r}$.*
3. $\sum_{i=1}^N \sum_{j=1}^N \langle \mathbf{v}_i, \mathbf{v}_j \rangle^2 = \frac{N^2}{r}$.

(The equivalence of the final condition is elementary but less obvious; the quantity on its left-hand side is sometimes called the *frame potential* [9].)

Definition 2.12. *The \mathbf{v}_i form an equiangular tight frame (ETF) if they form a UNTF, and there exists $\alpha \in [0, 1]$, called the coherence of the ETF, such that whenever $i \neq j$ then $|X_{ij}| = \alpha$.*

The following remarkable result shows that ETFs are extremal among UNTFs in the sense of *worst-case coherence*. Moreover, when an ETF exists, α is determined by N and r .

Proposition 2.13 (Welch Bound [70]). *If $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$ with $\|\mathbf{v}_i\|_2 = 1$, then*

$$\max_{\substack{1 \leq i, j \leq N \\ i \neq j}} |\langle \mathbf{v}_i, \mathbf{v}_j \rangle| \geq \sqrt{\frac{N - r}{r(N - 1)}}, \quad (11)$$

with equality if and only if $\mathbf{v}_1, \dots, \mathbf{v}_N$ form an ETF.

ETFs usually arise from combinatorial constructions and should generally be understood as rigid and highly structured objects. For instance, there remain many open problems about the pairs of dimensions (N, r) for which ETFs do or do not exist. More comprehensive references on these aspects of the theory of ETFs include [67, 16, 27].

2.2 Gramian Description of \mathcal{E}_4^N

Rather than directly working with the pseudomoment interpretation, we will study \mathcal{E}_4^N by pursuing an analogy with the following geometric description of the ellipsope:

$$\mathcal{E}_2^N = \{ \mathbf{X} \in \mathbb{R}_{\text{sym}}^{N \times N} : \exists \mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r \text{ such that } \|\mathbf{v}_i\|_2 = 1 \text{ and } \mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \}, \quad (12)$$

Besides being used in many geometric arguments about the ellipsope, which we will review later, in applications this description is central to the rounding procedures of [29, 57] as well as the efficient rank-constrained approximations of [13]. Our first result gives an analogous description of \mathcal{E}_4^N as Gram matrices of certain collections of unit vectors. The following family of matrices plays an important role in this description.

Definition 2.14. For $\mathbf{M} \in \mathbb{R}^{rN \times rN}$, we write $\mathbf{M}_{[ij]}$ with $i, j \in [N]$ for the $r \times r$ block in position (i, j) when \mathbf{M} is viewed as a block matrix. With this notation, let $\mathcal{B}(N, r) \subset \mathbb{R}_{\text{sym}}^{rN \times rN}$ consist of matrices \mathbf{M} satisfying the following properties:

1. $\mathbf{M} \succeq 0$.
2. $\mathbf{M}_{[ii]} = \mathbf{I}_r$ for all $i \in [N]$.
3. $\mathbf{M}_{[ij]} = \mathbf{M}_{[ij]}^\top$ for all $i, j \in [N]$.

In terms of these matrices, \mathcal{E}_4^N admits the following description.

Theorem 2.15. Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$, let $\mathbf{X} := \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathbb{R}_{\text{sym}}^{N \times N}$, let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have $\mathbf{v}_1, \dots, \mathbf{v}_N$ as its columns, and let $\mathbf{v} := \text{vec}(\mathbf{V}) \in \mathbb{R}^{rN}$ be the concatenation of the \mathbf{v}_i . Then, $\mathbf{X} \in \mathcal{E}_4^N$ if and only if $\sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N$ and there exists $\mathbf{M} \in \mathcal{B}(N, r)$ such that $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$.

Moreover, if $\mathbf{X} \in \mathcal{E}_4^N$ and a degree 4 pseudomoment matrix $\mathbf{Y} \in \mathbb{R}_{\text{sym}}^{N^2 \times N^2}$ extends \mathbf{X} , then there exists $\mathbf{M} \in \mathcal{B}(N, r)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$ and

$$\mathbf{Y} = (\mathbf{I}_N \otimes \mathbf{V})^\top \mathbf{M} (\mathbf{I}_N \otimes \mathbf{V}), \quad \text{i.e.} \quad (13)$$

$$Y_{(ij)(kl)} = \mathbf{v}_i^\top \mathbf{M}_{[jk]} \mathbf{v}_l \text{ for all } i, j, k, l \in [N]. \quad (14)$$

Conversely, if $\sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N$ and $\mathbf{M} \in \mathcal{B}(N, r)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$, then \mathbf{Y} as defined by (13) is a degree 4 pseudomoment matrix extending \mathbf{X} .

We think of \mathbf{M} as a *witness* of the fact that $\mathbf{X} \in \mathcal{E}_4^N$, an alternative to the pseudomoment witness \mathbf{Y} described in Proposition 2.4. The second, more detailed part of Theorem 2.15 gives one direction of the equivalence between these two types of witness; the other direction will be described in the course of the proof in Section 4.

2.3 Constraints on Pseudomoment Extensions

Through Theorem 2.15, we will next connect the structure of degree 4 pseudomoment extensions of $\mathbf{X} \in \mathcal{E}_2^N$ and the local geometry of \mathcal{E}_2^N near \mathbf{X} . Theorem 2.15 describes the membership of $\text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ in \mathcal{E}_4^N in terms of a semidefinite program, whose variable is $\mathbf{M} \in \mathcal{B}(N, r)$. Studying the dual of this semidefinite program and arguing through complementary slackness, we find that the optimal \mathbf{M} is highly constrained, as follows.

Lemma 2.16. *Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{S}^{r-1}$ be a spanning set, let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns, let $\mathbf{v} := \text{vec}(\mathbf{V}) \in \mathbb{R}^{rN}$ be the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N$, let $\mathbf{X} := \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathcal{E}_2^N$, and let $\mathbf{M}^* \in \mathcal{B}(N, r)$ be such that $\mathbf{v}^\top \mathbf{M}^* \mathbf{v} = N^2$.*

Then, all eigenvectors of \mathbf{M}^ with nonzero eigenvalue belong to the subspace*

$$V_{\text{sym}} := \{ \text{vec}(\mathbf{S}\mathbf{V}) : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r} \} \subset \mathbb{R}^{rN}. \quad (15)$$

Additionally, \mathbf{v} is an eigenvector of \mathbf{M}^ with eigenvalue N , and all eigenvectors of \mathbf{M}^* with nonzero eigenvalue that are orthogonal to \mathbf{v} belong to the subspace*

$$V'_{\text{sym}} := \left\{ \text{vec}(\mathbf{S}\mathbf{V}) : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}, \mathbf{v}_i^\top \mathbf{S} \mathbf{v}_i = 0 \text{ for } i \in [N] \right\} \subset \mathbb{R}^{rN}. \quad (16)$$

We next apply (13), which shows how a spectral decomposition of \mathbf{M}^* gives an expression for the associated pseudomoment matrix \mathbf{Y} as a sum of $\text{rank}(\mathbf{M}^*)$ (not necessarily orthogonal) rank one matrices, which are constrained by Lemma 2.16. It turns out that these latter constraints are similar to those appearing in results of [53, 49] connecting the smallest face of \mathcal{E}_2^N containing \mathbf{X} to $\text{span}(\{\mathbf{v}_i \mathbf{v}_i^\top\}_{i=1}^N)$, which lets us describe the constraints on \mathbf{Y} concisely in terms of the local geometry of \mathcal{E}_2^N near \mathbf{X} .

Theorem 2.17. *Suppose $\mathbf{X} \in \mathcal{E}_4^N$ and \mathbf{Y} is a degree 4 pseudomoment matrix extending \mathbf{X} . Then, $\mathbf{Y} \succeq \text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top$, and all eigenvectors of $\mathbf{Y} - \text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top$ with nonzero eigenvalue belong to the subspace $\text{vec}(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}))$.*

Consequently,

$$\text{rank}(\mathbf{Y}) \leq \dim \left(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}) \right) + 1 \quad (17)$$

$$= \frac{\text{rank}(\mathbf{X})(\text{rank}(\mathbf{X}) + 1)}{2} - \text{rank}(\mathbf{X}^{\odot 2}) + 1 \quad (18)$$

$$\leq \frac{\text{rank}(\mathbf{X})(\text{rank}(\mathbf{X}) + 1)}{2}. \quad (19)$$

In particular, if \mathbf{X} is an extreme point of \mathcal{E}_2^N and is extendable to a degree 4 pseudomoment matrix \mathbf{Y} , then $\text{rank}(\mathbf{Y}) = \text{rank}(\mathbf{X}) = 1$, and $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ and $\mathbf{Y} = (\mathbf{x} \otimes \mathbf{x})(\mathbf{x} \otimes \mathbf{x})^\top$ for some $\mathbf{x} \in \{\pm 1\}^N$.

Remark 2.18. *The matrix $\mathbf{Y} - \text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top$ is quite natural in the pseudomoment framework: entry $(ij)(k\ell)$ of this matrix contains the difference between the pseudoexpectation of $x_i x_j x_k x_\ell$ and the product of the pseudoexpectations of $x_i x_j$ and $x_k x_\ell$. It is natural to think of this quantity as the pseudocovariance of $x_i x_j$ and $x_k x_\ell$, and it is then not surprising that the SOS constraints imply that the pseudocovariance matrix is psd. We are not aware, however, of previous results on SOS optimization that make direct use of the pseudocovariance matrix.*

(As we will discuss in Section 5, the equality (18) is a previously known result of [53, 49]. Recall also that $\text{pert}_{\mathcal{E}_2^N}(\mathbf{X})$ as a subspace has the same dimension as $\text{face}_{\mathcal{E}_2^N}(\mathbf{X})$ as a convex set.) The

final claim gives a strong, albeit non-quantitative, suggestion that \mathcal{E}_4^N is a substantially tighter relaxation of \mathcal{C}^N than \mathcal{E}_2^N : it implies that no “spurious” extreme points of \mathcal{E}_2^N that are not already extreme points of \mathcal{C}^N persist after constraining to \mathcal{E}_4^N .

The bounds (18) and (19) are similar in form to the Pataki bound on the rank of extreme points of feasible sets of general SDPs [59]. Because of the very large number of linear constraints in SDPs arising from SOS optimization, however, the Pataki bound is less effective in this setting. In particular, the Pataki bound gives, at best,

$$\text{rank}(\mathbf{Y}) \leq (1 + o_{N \rightarrow \infty}(1))\sqrt{2m} \quad (20)$$

for \mathbf{Y} an extreme point of the set of degree 4 pseudomoment matrices, where m is the number of linear constraints in the definition of an $N^2 \times N^2$ degree 4 pseudomoment matrix. The degree 4 SOS constraints require that for each subset $\{i, j, k, \ell\} \subseteq [N]$, the permutation invariance of $Y_{(ij)(k\ell)}$ be enforced. There are $\binom{N}{4} \sim \frac{1}{24}N^4$ such subsets and 24 “copies” whose equality must be enforced for each, of which it suffices to consider 12 since the constraint that the matrix \mathbf{Y} be symmetric is not counted. Thus $m = (1 - o_{N \rightarrow \infty}(1))\frac{11}{24}N^4$, whereby the right-hand side of (20) is $(1 - o_{N \rightarrow \infty}(1))\sqrt{\frac{11}{12}}N^2$.

On the other hand, even with the weaker of our bounds (19) and the naive further bound $\text{rank}(\mathbf{X}) \leq N$, we obtain the stronger claim that *any* degree 4 pseudomoment matrix, not necessarily an extreme point, has rank at most $\binom{N+1}{2} \sim \frac{1}{2}N^2$. Additionally, the stronger inequality (18) is tight, achieved for instance by the degree 4 pseudomoment matrix \mathbf{Y} where $Y_{(ij)(k\ell)} = 1$ if each index i, j, k, ℓ appears an even number of times and $Y_{(ij)(k\ell)} = 0$ otherwise (this matrix is the average of $(\mathbf{x} \otimes \mathbf{x})(\mathbf{x} \otimes \mathbf{x})^\top$ over all $\mathbf{x} \in \{\pm 1\}^N$), which extends $\mathbf{X} = \mathbf{I}_N$. Our bound would also give improved control of $\text{rank}(\mathbf{Y})$ in the case where \mathbf{X} is low-rank (say, when $\text{rank}(\mathbf{X}) \leq \delta N$ for small enough δ). The question of whether, in fact, the degree 2 pseudomoments of an optimal degree 4 pseudomoment matrix are typically low-rank in practical or random problems appears not to have been studied extensively and would be an interesting question for future work. The result (17) casts some doubt on this natural conjecture, as it implies that there is less “room” to build a degree 4 pseudomoment matrix extending a degree 2 pseudomoment matrix that lies on a low-dimensional face of \mathcal{E}_2^N .

2.4 Examples from Equiangular Tight Frames

We next analyze the highly structured case of ETF Gram matrices, where the constraints of the previous section may guide the search for a degree 4 pseudomoment matrix extending a given degree 2 pseudomoment matrix. The main reason that it is convenient to work with ETFs is that, when \mathbf{X} is the Gram matrix of an ETF, then $|X_{ij}|$ takes only two values, 1 when $i = j$ and some $\alpha \in [0, 1]$ otherwise. Therefore, in particular, the matrix $\mathbf{X}^{\odot 2}$ is very simple,

$$\mathbf{X}^{\odot 2} = (1 - \alpha^2)\mathbf{I}_N + \alpha^2\mathbf{1}_N\mathbf{1}_N^\top. \quad (21)$$

As we have seen in Theorem 2.17, $\mathbf{X}^{\odot 2}$ is intimately related to $\text{pert}_{\mathcal{E}_2^N}(\mathbf{X})$ and therefore to the possible degree 4 pseudomoment extensions of \mathbf{X} . In the case of ETFs, its simple structure makes it possible to compute an explicit (albeit naive) guess for a degree 4 pseudomoment extension, which rather surprisingly turns out to be correct.

By such reasoning, we obtain a complete characterization of membership in \mathcal{E}_4^N for ETF Gram matrices \mathbf{X} , which is quite simple in that it depends only on the dimension and rank of \mathbf{X} . This result is as follows.

Theorem 2.19. Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$ form an ETF, and let $\mathbf{X} := \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$. Then, $\mathbf{X} \in \mathcal{E}_4^N$ if and only if $N < \frac{r(r+1)}{2}$ or $r = 1$. If $r = 1$, then $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ for $\mathbf{x} \in \{\pm 1\}^N$, and a degree 4 pseudomoment matrix \mathbf{Y} extending \mathbf{X} is given by $\mathbf{Y} = (\mathbf{x} \otimes \mathbf{x})(\mathbf{x} \otimes \mathbf{x})^\top$. If $r > 1$ and $N < \frac{r(r+1)}{2}$, then, letting $\mathbf{P}_{\text{vec}(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}))}$ be the orthogonal projection matrix to $\text{vec}(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X})) \subset \mathbb{R}^{N^2}$, a degree 4 pseudomoment matrix \mathbf{Y} extending \mathbf{X} is given by

$$\mathbf{Y} = \text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top + \frac{N^2(1 - \frac{1}{r})}{\frac{r(r+1)}{2} - N} \mathbf{P}_{\text{vec}(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}))}, \text{ i.e.} \quad (22)$$

$$Y_{(ij)(kl)} = \frac{\frac{r(r-1)}{2}}{\frac{r(r+1)}{2} - N} (X_{ij}X_{kl} + X_{ik}X_{jl} + X_{il}X_{jk}) - \frac{r^2(1 - \frac{1}{N})}{\frac{r(r+1)}{2} - N} \sum_{m=1}^N X_{im}X_{jm}X_{km}X_{lm}. \quad (23)$$

As we will discuss further in the sequel, it is always the case that $N \leq \frac{r(r+1)}{2}$ for an ETF (this is the Gerzon bound [51]), and the maximal ETFs with $N = \frac{r(r+1)}{2}$ are notoriously elusive combinatorial objects; for instance, they are known to exist for only four values of N , and the question of their existence is open for infinitely many values of N [27]. Our result invokes another regard in which these ETFs are extremal, which was in fact present but perhaps unnoticed in existing results (in particular in an elegant proof of the Gerzon bound that we will present later): maximal ETF Gram matrices are the only ETF Gram matrices that are extreme points of \mathcal{E}_2^N (thus, by Theorem 2.17, these Gram matrices cannot belong to \mathcal{E}_4^N).

In our argument it will become clear that the case of ETFs (those non-maximal ones that do belong to \mathcal{E}_4^N) is perhaps the simplest possible situation for degree 4 pseudomoments over the hypercube: as shown in (22), the degree 4 pseudomoment matrix \mathbf{Y} will have only two distinct positive eigenvalues, and will equal of the sum of the rank one matrix $\text{vec}(\mathbf{X})\text{vec}(\mathbf{X})^\top$, which contributes the “naive” pseudomoment value $X_{ij}X_{kl}$, with a constant multiple of the projection matrix onto the subspace $\text{vec}(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}))$, which contributes the remaining “symmetrization” term appearing in (23).

2.5 Applications

New inequalities in \mathcal{E}_4^N . There are many results in combinatorial optimization enumerating linear inequalities satisfied by \mathcal{C}^N (see e.g. [22]). The practical purpose of this pursuit is that such linear inequalities may be included in linear programming (LP) relaxations of \mathcal{C}^N , which are typically more efficient than the SDP relaxations we work with here. The putative convenience of SDP relaxations is that they do not require their user to know specifically which inequalities will be relevant for a given problem; the psd constraint captures many relevant inequalities at once. For theoretical understanding, however, it is again important to know which specific inequalities over \mathcal{C}^N are satisfied at which degrees of SOS relaxation, since those inequalities may then be used as analytical tools.

Yet, to the best of our knowledge, very few inequalities over \mathcal{C}^N are known to be satisfied in \mathcal{E}_4^N but not \mathcal{E}_2^N ; indeed, it appears that the only infinite such family known before this work was the *triangle inequalities*,

$$-s_i s_j X_{ij} - s_j s_k X_{jk} - s_i s_k X_{ik} \leq 1 \text{ for } \mathbf{X} \in \mathcal{E}_4^N, \mathbf{s} \in \{\pm 1\}^N. \quad (24)$$

Guided by the results from the previous section, we find a new family of similar but independent inequalities. First, from the negative result of Theorem 2.19, we obtain concrete examples of matrices $\mathbf{X} \in \mathcal{E}_2^N \setminus \mathcal{E}_4^N$, namely the Gram matrices of ETFs with $N = \frac{r(r+1)}{2}$. As mentioned

before, these are only known to exist for four specific dimensions, namely $r \in \{2, 3, 7, 23\}$. By convex duality, there must exist certificates that these matrices do not belong to \mathcal{E}_4^N , taking the form of linear inequalities that hold over \mathcal{E}_4^N but fail to hold for these matrices. Indeed, for the smallest two examples $r \in \{2, 3\}$, a triangle inequality is a valid certificate of infeasibility.

For $r = 7$, on the other hand, the absolute value of the off-diagonal entries of the Gram matrix is $\alpha = \frac{1}{3}$, so the triangle inequalities are satisfied, and the certificates of infeasibility must be new inequalities which cannot be obtained as linear combinations of triangle inequalities. We compute these certificates numerically and identify the constants that arise by hand to allow the certificates to be validated by symbolic computation (this amounts to checking that a certain $N^2 \times N^2$ matrix is psd, where in this case $N^2 = 28^2 = 784$).

For $r = 23$ the same appears to occur numerically and a similar argument shows that yet another independent family of inequalities must arise as the certificates of infeasibility, but the symbolic verification of such a certificate is a much larger problem which a naive software implementation does not solve in a reasonable time. We thus only present the verified result for $r = 7$ here as a proof of concept, leaving both further computational verification of exact inequalities and further theoretical analysis of these certificates to future work.

Theorem 2.20. *Let \mathbf{Z} be the Gram matrix of an ETF of 28 vectors in \mathbb{R}^7 . Then, for any $\mathbf{X} \in \mathcal{E}_4^N$ and any $\pi : [28] \rightarrow [N]$ injective,*

$$\sum_{1 \leq i < j \leq 28} \operatorname{sgn}(Z_{ij}) X_{\pi(i)\pi(j)} \leq 112, \quad (25)$$

and this inequality cannot be obtained as a linear combination of the triangle inequalities

$$-s_i s_j X_{ij} - s_j s_k X_{jk} - s_i s_k X_{ik} \leq 1 \text{ for } \mathbf{s} \in \{\pm 1\}^N. \quad (26)$$

As we will detail in the sequel, there is a general correspondence between ETFs and strongly regular graphs (SRGs) [28], under which the ETF of 28 vectors in \mathbb{R}^7 (which is unique up to negating a subset of its vectors) corresponds to the Schläfli graph, a remarkably symmetrical 16-regular graph on 27 vertices (see e.g. [14, 18] for examples of its structure and significance in combinatorics). We thus refer to these inequalities as *Schläfli inequalities*.

As a point of comparison, since $\mathbf{Z} \in \mathcal{E}_2^N$, the half-space parallel to that defined by (25) most tightly bounding \mathcal{E}_2^N must have the right-hand side at least

$$\sum_{1 \leq i < j \leq 28} \operatorname{sgn}(Z_{ij}) Z_{ij} = \sum_{1 \leq i < j \leq 28} |Z_{ij}| = \frac{28(28-1)}{2} \cdot \frac{1}{3} = 126 > 112. \quad (27)$$

Thus, the Schläfli inequalities describe directions in the vector space of symmetric matrices along which \mathcal{E}_4^N is strictly “narrower” than \mathcal{E}_2^N .

New proofs of complexity of parity inequalities. Lastly, our results on ETFs imply a special case of the following theorem of Laurent, which shows a lower bound for what degree of the SOS hierarchy is required to describe the cut polytope exactly. (In the same work this bound was conjectured to be optimal, which was later proven in [24].)

Proposition 2.21 (Theorems 5 and 6 of [46]). *Let $N \geq 3$ be odd. Define*

$$\mathbf{X}^{(N)} := \left(1 + \frac{1}{N-1}\right) \mathbf{I}_N - \frac{1}{N-1} \mathbf{1}\mathbf{1}^\top. \quad (28)$$

Then, $\mathbf{X}^{(N)} \in \mathcal{E}_{N-1}^N \setminus \mathcal{C}^N$, and the degree $N-1$ pseudomoment matrix $\mathbf{Z}^{(N)} \in \mathbb{R}^{[N]^{(N-1)/2} \times [N]^{(N-1)/2}}$ whose entries are given by

$$Z_{\mathbf{st}}^{(N)} := (-1)^{|\text{odd}(\mathbf{sot})|/2} \prod_{\substack{i \text{ odd} \\ 1 \leq i < |\text{odd}(\mathbf{sot})|}} \frac{i}{N-i} \quad (29)$$

extends $\mathbf{X}^{(N)}$. In particular, $\mathcal{E}_{N-1}^N \neq \mathcal{C}^N$.

The ‘‘parity inequality’’ we refer to in the heading of this section is the fact that $\mathbf{X}^{(N)} \notin \mathcal{C}^N$, which follows from the fact that when N is odd, then $\mathbf{1}_N^\top \mathbf{X} \mathbf{1}_N \geq 1$ for all $\mathbf{X} \in \mathcal{C}^N$ (which in turn follows from the fact that when $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ with $\mathbf{x} \in \{\pm 1\}^N$, then this simply says $(\sum_{i=1}^N x_i)^2 \geq 1$), while $\mathbf{1}_N^\top \mathbf{X}^{(N)} \mathbf{1}_N = 0$. Proposition 2.21 then says that this parity inequality fails to hold over \mathcal{E}_{N-1}^N .

Observing that $\mathbf{X}^{(N)}$ is the Gram matrix of an ETF, we are able to reproduce a weaker version of this result as a corollary of Theorem 2.19, where \mathcal{E}_{N-1}^N is replaced with \mathcal{E}_4^N .

Corollary 2.22. *Let $N \geq 4$. Then, the matrix $\mathbf{Y} \in \mathbb{R}^{N^2 \times N^2}$ defined by*

$$Y_{(ij)(kl)}^{(N)} = \begin{cases} 1 & : |\text{odd}((ijkl))| = 0, \\ -\frac{1}{N-3} & : |\text{odd}((ijkl))| = 2, \\ \frac{3}{(N-1)(N-3)} & : |\text{odd}((ijkl))| = 4 \end{cases} \quad (30)$$

is a degree 4 pseudomoment matrix extending $\mathbf{X}^{(N)}$.

Though this result is weaker than the full Proposition 2.21, the technique of its proof is far simpler: we do not rely on any of the theory of association schemes or hypergeometric series used in the argument of [46], and we obtain the correct value for the degree 4 pseudomoments, the key result that $\mathbf{Y} \succeq \mathbf{0}$, and explicit descriptions of the eigenvalues and eigenvectors of \mathbf{Y} (via the matrix formula (22)) from a single straightforward linear algebra calculation. We believe it is likely that a generalization of the methods presented here can yield the full result of Laurent’s theorem in a similarly simplified fashion.

3 Notations

Boldface uppercase letters (\mathbf{X}) denote matrices, and boldface lowercase letters (\mathbf{x}) denote vectors. Entries of matrices and vectors are written without boldface and with subscripts (X_{ij} , x_i), except if the vector or matrix itself has a subscript, in which case parentheses are used ($(\mathbf{x}_i)_j$, $(\mathbf{X}_i)_{jk}$).

\mathbf{I}_d denotes the $d \times d$ identity matrix, $\mathbf{1}_d$ denotes the all-ones vector of length d , and $\hat{\mathbf{1}}_d$ denotes its normalization to a unit vector, $\hat{\mathbf{1}}_d = \frac{1}{\sqrt{d}} \mathbf{1}_d$. The subscripts from these notations will be omitted when the suitable dimension is clear from context.

$S^{r-1} \subset \mathbb{R}^r$ denotes the sphere of unit radius in \mathbb{R}^r . $\mathbb{R}^{k \times k}$ (resp. $\mathbb{R}_{\text{sym}}^{k \times k}$, $\mathbb{R}_{\text{antisym}}^{k \times k}$) denotes the set of $k \times k$ matrices (resp. symmetric, antisymmetric $k \times k$ matrices) with real entries. $\mathcal{O}(k)$ denotes the group of $k \times k$ orthogonal matrices.

For a matrix \mathbf{V} , $\text{row}(\mathbf{V})$ denotes the span of its rows, $\ker(\mathbf{V})$ denotes its kernel as an operator (also the orthogonal complement of $\text{row}(\mathbf{V})$), $\text{rank}(\mathbf{V})$ denotes its rank, and $\text{null}(\mathbf{V})$ denotes its nullity, the dimension of its kernel. For a vector $\mathbf{v} \in \mathbb{R}^k$, $\text{diag}(\mathbf{v}) \in \mathbb{R}^{k \times k}$ is the diagonal matrix with $\text{diag}(\mathbf{v})_{ii} = v_i$. For matrices \mathbf{X}, \mathbf{Y} of the same dimensions, $\mathbf{X} \odot \mathbf{Y}$ denotes the Hadamard or entrywise product, and $\mathbf{X}^{\odot k}$ denotes the matrix obtained by taking the k th power of each entry

of \mathbf{X} . For any subspace $V \subseteq \mathbb{R}^r$, $\mathbf{P}_V \in \mathbb{R}^{r \times r}$ denotes the orthogonal projection matrix to V . For $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^k$, $\text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathbb{R}^{N \times N}$ denotes the Gram matrix $(\langle \mathbf{v}_i, \mathbf{v}_j \rangle)_{i,j=1}^N$.

We will work extensively with block matrices, for which we propose a non-standard but helpful notation. When $\mathbf{X} \in \mathbb{R}^{ab \times cd}$ and we have declared \mathbf{X} to be a $b \times d$ matrix of blocks of size $a \times c$, then $\mathbf{X}_{[ij]} \in \mathbb{R}^{a \times c}$ denotes the block in position $(i, j) \in [b] \times [d]$. Similarly, when $\mathbf{x} \in \mathbb{R}^{ab}$ and we have declared \mathbf{x} to be divided into blocks of length a , then $\mathbf{x}_{[i]} \in \mathbb{R}^a$ denotes the block in position $i \in [b]$. When $c = a$, so that the blocks of \mathbf{X} are square, we define the *partial transpose* operation \mathbf{X}^Γ to produce the matrix where every $a \times a$ block of \mathbf{X} is transposed.²

4 Gramian Description of \mathcal{E}_4^N : Theorem 2.15

In this section we give the proof of Theorem 2.15, followed by some ancillary results providing intuitive interpretations of its statement and suggesting some connections to the notion of *separability* as studied in quantum information theory. Recall that Theorem 2.15 describes an equivalence between the pseudomoment witness $\mathbf{Y} \in \mathbb{R}^{N^2 \times N^2}$ extending $\mathbf{X} \in \mathcal{E}_2^N$ and the Gram vector witness $\mathbf{M} \in \mathcal{B}(N, r)$. The basic idea is that \mathbf{M} describes certain rotations by which the blocks of \mathbf{Y} must be related because of the pseudomoment matrix constraints. We will show first how to build \mathbf{M} from \mathbf{Y} , and then how to build \mathbf{Y} from \mathbf{M} .

4.1 Proof of Theorem 2.15: Pseudomoment Witness to Gram Vector Witness

Let $\mathbf{Y} \in \mathbb{R}^{N^2 \times N^2}$ be a degree 4 pseudomoment matrix extending $\mathbf{X} \in \mathbb{R}^{N \times N}$, where for some $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$, $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$. Let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns, and let $\mathbf{v} = \text{vec}(\mathbf{V}) \in \mathbb{R}^{rN}$ be the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N$. We will then show that there exists $\mathbf{M} \in \mathcal{B}(N, r)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$ and

$$\mathbf{Y} = (\mathbf{I}_N \otimes \mathbf{V})^\top \mathbf{M} (\mathbf{I}_N \otimes \mathbf{V}). \quad (31)$$

We first analyze the special case $r = \text{rank}(\mathbf{X})$, then extend to the general case.

Case 1: $r = \text{rank}(\mathbf{X})$. We build \mathbf{M} based on a suitable factorization of \mathbf{Y} . Let $r' := \text{rank}(\mathbf{Y}) \geq r$, then there exists $\mathbf{A} \in \mathbb{R}^{r' \times N^2}$ such that $\mathbf{Y} = \mathbf{A}^\top \mathbf{A}$. Let us expand in blocks

$$\mathbf{A} = [\mathbf{A}_1 \quad \mathbf{A}_2 \quad \cdots \quad \mathbf{A}_N], \quad (32)$$

for $\mathbf{A}_i \in \mathbb{R}^{r' \times N}$. Since $\mathbf{A}_1^\top \mathbf{A}_1 = \mathbf{Y}_{[11]} = \mathbf{X} = \mathbf{V}^\top \mathbf{V}$, there exists $\mathbf{Z} \in \mathbb{R}^{r' \times r}$ such that $\mathbf{A}_1 = \mathbf{Z} \mathbf{V}$ and $\mathbf{Z}^\top \mathbf{Z} = \mathbf{I}_r$. By adding extra columns, we may extend \mathbf{Z} to an orthogonal matrix $\tilde{\mathbf{Z}} \in \mathcal{O}(r')$. The factorization $\mathbf{Y} = \mathbf{A}^\top \mathbf{A}$ is unchanged by multiplying \mathbf{A} on the left by any element of $\mathcal{O}(r')$. By performing this transformation with $\tilde{\mathbf{Z}}$, we may assume without loss of generality that \mathbf{A} is chosen such that

$$\mathbf{A}_1 = \left[\begin{array}{c} \mathbf{V} \\ \mathbf{0} \end{array} \right] \left. \begin{array}{l} \} r \\ \} r' - r \end{array} \right. \quad (33)$$

where the numbers following the braces show the dimensionality of the matrix blocks.

Now, since $\mathbf{A}_i^\top \mathbf{A}_i = \mathbf{Y}_{[ii]} = \mathbf{X} = \mathbf{A}_1^\top \mathbf{A}_1$ for every $i \in [N]$ (since, by the degree 4 pseudomoment conditions, $Y_{(ik)(il)} = Y_{(ii)(kl)} = Y_{(11)(kl)} = X_{kl}$), there must exist $\mathbf{I}_{r'} = \mathbf{Q}_1, \dots, \mathbf{Q}_N \in \mathcal{O}(r')$ such

²This notation, standard in the quantum information literature, is justified by the symbol Γ being ‘‘half of’’ the transpose symbol \top .

that $\mathbf{A}_i = \mathbf{Q}_i \mathbf{A}_1$. Let us expand \mathbf{Q}_i in blocks,

$$\mathbf{Q}_i = \begin{bmatrix} \underbrace{\mathbf{U}_i}_r & \underbrace{\tilde{\mathbf{U}}_i}_{r'-r} \end{bmatrix}. \quad (34)$$

We then have

$$\mathbf{A}_i = \mathbf{Q}_i \mathbf{A}_1 = \mathbf{U}_i \mathbf{V}. \quad (35)$$

(The extra variable $\tilde{\mathbf{U}}_i$ will not be used in the argument.) Therefore, the blocks of \mathbf{Y} are given by

$$\mathbf{Y}_{[ij]} = \mathbf{A}_i^\top \mathbf{A}_j = \mathbf{V}^\top \mathbf{U}_i^\top \mathbf{U}_j \mathbf{V}. \quad (36)$$

By the permutation symmetry of \mathbf{Y} , every such block is symmetric. Since \mathbf{V} has full rank, $\mathbf{V}\mathbf{V}^\top$ is invertible, and therefore the matrix $(\mathbf{V}\mathbf{V}^\top)^{-1} \mathbf{V}\mathbf{Y}_{[ij]}\mathbf{V}^\top (\mathbf{V}\mathbf{V}^\top)^{-1} = \mathbf{U}_i^\top \mathbf{U}_j$ is also symmetric.

We now define \mathbf{M} blockwise by

$$\mathbf{M}_{[ij]} := \mathbf{U}_i^\top \mathbf{U}_j. \quad (37)$$

Then $\mathbf{M} \succeq \mathbf{0}$ by construction, $\mathbf{M}_{[ii]} = \mathbf{I}_r$ since this is the upper left block of $\mathbf{Q}_i^\top \mathbf{Q}_i = \mathbf{I}_{r'}$, and $\mathbf{M}_{[ij]}$ is symmetric by the preceding derivation. Thus, $\mathbf{M} \in \mathcal{B}(N, r)$. By (36), we also have

$$\mathbf{Y} = (\mathbf{I}_N \otimes \mathbf{V})^\top \mathbf{M} (\mathbf{I}_N \otimes \mathbf{V}). \quad (38)$$

It remains only to check that $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$:

$$\mathbf{v}^\top \mathbf{M} \mathbf{v} = \sum_{i=1}^N \sum_{j=1}^N \mathbf{v}_i^\top \mathbf{M}_{[ij]} \mathbf{v}_j = \sum_{i=1}^N \sum_{j=1}^N (\mathbf{V}^\top \mathbf{U}_i^\top \mathbf{U}_j \mathbf{V})_{ij} = \sum_{i=1}^N \sum_{j=1}^N Y_{(ii)(jj)} = N^2, \quad (39)$$

completing the proof of the first case.

Case 2: $r > \text{rank}(\mathbf{X})$. We will reduce this case to the previous case. Let $r_0 = \text{rank}(\mathbf{X}) < r$. Fix Gram vectors $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^{r_0}$ such that $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$, and, by the previous argument, choose $\mathbf{M} \in \mathcal{B}(N, r_0)$ having $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$.

Suppose that $\mathbf{v}'_1, \dots, \mathbf{v}'_N \in \mathbb{R}^r$ such that $\mathbf{X} = \text{Gram}(\mathbf{v}'_1, \dots, \mathbf{v}'_N)$. Let \mathbf{v}' be the concatenation of $\mathbf{v}'_1, \dots, \mathbf{v}'_N$. Since the Gram matrices of $\mathbf{v}_1, \dots, \mathbf{v}_N$ and $\mathbf{v}'_1, \dots, \mathbf{v}'_N$ are equal, there must exist $\mathbf{Z} \in \mathbb{R}^{r \times r_0}$ with $\mathbf{Z}\mathbf{v}_i = \mathbf{v}'_i$ for each $i \in [N]$ and $\mathbf{Z}^\top \mathbf{Z} = \mathbf{I}_{r_0}$. Define $\mathbf{M}' \in \mathbb{R}^{rN \times rN}$ to have blocks

$$\mathbf{M}'_{[ij]} := \begin{cases} \mathbf{Z}\mathbf{M}_{[ij]}\mathbf{Z}^\top & : i \neq j, \\ \mathbf{I}_r & : i = j. \end{cases} \quad (40)$$

Equivalently,

$$\mathbf{M}' = (\mathbf{I}_N \otimes \mathbf{Z})\mathbf{M}(\mathbf{I}_N \otimes \mathbf{Z})^\top + \mathbf{I}_N \otimes (\mathbf{I}_r - \mathbf{Z}\mathbf{Z}^\top). \quad (41)$$

Since $\mathbf{Z}\mathbf{Z}^\top \preceq \mathbf{I}_r$ (the left-hand side is a projection matrix), $\mathbf{M}' \succeq \mathbf{0}$, and by construction $\mathbf{M}'_{[ii]} = \mathbf{I}_r$ and $\mathbf{M}'_{[ij]}$ is symmetric. Thus, $\mathbf{M}' \in \mathcal{B}(N, r)$.

We also have

$$\begin{aligned} \mathbf{v}'^\top \mathbf{M}' \mathbf{v}' &= \sum_{i=1}^N \|\mathbf{v}'_i\|_2^2 + \sum_{\substack{1 \leq i, j \leq N \\ i \neq j}} \mathbf{v}'_i^\top \mathbf{M}'_{[ij]} \mathbf{v}'_j \\ &= N + \sum_{\substack{1 \leq i, j \leq N \\ i \neq j}} \mathbf{v}_i^\top \mathbf{M}_{[ij]} \mathbf{v}_j \\ &= N^2. \end{aligned} \quad (42)$$

Lastly, we check the formula for the entries of \mathbf{Y} , now distinguishing the cases $i = j$ and $i \neq j$:

$$Y_{(ii)(k\ell)} = X_{k\ell} = \langle \mathbf{v}'_k, \mathbf{v}'_\ell \rangle = \mathbf{v}'_k{}^\top \mathbf{M}'_{[ii]} \mathbf{v}'_\ell, \quad (43)$$

$$\begin{aligned} Y_{(ij)(k\ell)} &= \mathbf{v}_k{}^\top \mathbf{M}_{[ij]} \mathbf{v}_\ell \\ &= \mathbf{v}'_k{}^\top \mathbf{Z} \mathbf{M}_{[ij]} \mathbf{Z}^\top \mathbf{v}'_\ell \\ &= \mathbf{v}'_k{}^\top \mathbf{M}'_{[ij]} \mathbf{v}'_\ell \text{ (for } i \neq j), \end{aligned} \quad (44)$$

completing the proof.

4.2 Proof of Theorem 2.15: Gram Vector Witness to Pseudomoment Witness

Suppose that $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathbb{R}^{N \times N}$ for some $\mathbf{v}_i \in \mathbb{R}^r$ having $\sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N$. Let \mathbf{v} be the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N$. Suppose also that $\mathbf{M} \in \mathcal{B}(N, r)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$. We will show that $\mathbf{Y} \in \mathbb{R}^{N^2 \times N^2}$ defined by

$$Y_{(ij)(k\ell)} = \mathbf{v}_i{}^\top \mathbf{M}_{[jk]} \mathbf{v}_\ell \quad (45)$$

is a degree 4 pseudomoment matrix. Recall that this requires the following properties to hold:

1. $\mathbf{Y} \succeq \mathbf{0}$.
2. $Y_{(ij)(kk)}$ does not depend on the index k .
3. $Y_{(ii)(ii)} = 1$ for every $i \in [N]$.
4. $Y_{(ij)(k\ell)}$ is invariant under permutations of the indices i, j, k, ℓ .

(That the upper left $N \times N$ block of \mathbf{Y} is \mathbf{X} follows from Property 4 and that $\mathbf{M}_{[ii]} = \mathbf{I}_r$.) We will obtain these one by one below. This essentially just entails reversing the derivation of the previous part; however, verifying some of the properties of \mathbf{Y} will require a more detailed understanding of the factorization of \mathbf{M} that we used.

The simplest is Property 1: since $\mathbf{M} \succeq \mathbf{0}$, there exist some $\mathbf{U}_1, \dots, \mathbf{U}_N \in \mathbb{R}^{r' \times r}$ for some $r' \geq 1$ such that $\mathbf{M}_{[jk]} = \mathbf{U}_j{}^\top \mathbf{U}_k$. Thus,

$$Y_{(ij)(k\ell)} = \mathbf{v}_i{}^\top \mathbf{U}_j{}^\top \mathbf{U}_k \mathbf{v}_\ell = \langle \mathbf{U}_j \mathbf{v}_i, \mathbf{U}_k \mathbf{v}_\ell \rangle, \quad (46)$$

so $\mathbf{Y} = \text{Gram}(\mathbf{U}_1 \mathbf{v}_1, \dots, \mathbf{U}_N \mathbf{v}_N) \succeq \mathbf{0}$.

For Properties 2 and 3, we will first need some basic results on the spectrum of $\mathbf{M} \in \mathcal{B}(N, r)$. The proofs of these facts are given in Appendix B.

Proposition 4.1. *Let $\mathbf{M} \in \mathcal{B}(N, r)$. Then,*

1. $\|\mathbf{M}_{[ij]}\| \leq 1$ for all $i, j \in [N]$;
2. $\|\mathbf{M}\| \leq N$;
3. if $\mathbf{M} \mathbf{v} = N \mathbf{v}$, and $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^{rN}$ is the concatenation of $\mathbf{v}_i \in \mathbb{R}^r$, then the norms $\|\mathbf{v}_i\|_2$ are all equal, and $\mathbf{M}_{[ij]} \mathbf{v}_j = \mathbf{v}_i$ for all $i, j \in [N]$.

From Claim 2 in the Proposition, since $\|\mathbf{v}\|_2^2 = \text{Tr}(\mathbf{X}) = N$, then if $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$ we must have $\mathbf{M} \mathbf{v} = N \mathbf{v}$. Therefore, by Claim 3, $\|\mathbf{v}_i\|_2 = 1$ for each $i \in [N]$. Also by Claim 3, we have

$$Y_{(ij)(kk)} = \mathbf{v}_i{}^\top \mathbf{M}_{[jk]} \mathbf{v}_k = \langle \mathbf{v}_i, \mathbf{v}_j \rangle. \quad (47)$$

This gives Property 2, and taking $i = j = k$ gives Property 3 since $\|\mathbf{v}_i\|_2 = 1$.

Property 4 is more subtle to establish. First, for a moment treating i, j, k, ℓ as merely four distinct symbols, note that the symmetric group on $\{i, j, k, \ell\}$ is generated by the three transpositions (ij) , (jk) , and $(k\ell)$. Therefore, to establish Property 4 it suffices to show the three equalities

$$Y_{(ij)(k\ell)} = Y_{(ji)(k\ell)} = Y_{(ij)(\ell k)} = Y_{(ik)(j\ell)} \quad (48)$$

for all $i, j, k, \ell \in [N]$. One equality follows directly from both $\mathbf{M}_{[jk]}$ and \mathbf{M} being symmetric, whereby $\mathbf{M}_{[jk]} = \mathbf{M}_{[kj]}$:

$$Y_{(ij)(k\ell)} = \mathbf{v}_i^\top \mathbf{M}_{[jk]} \mathbf{v}_\ell = \mathbf{v}_i^\top \mathbf{M}_{[kj]} \mathbf{v}_\ell = Y_{(ik)(j\ell)}. \quad (49)$$

The others require a more detailed argument involving a factorization of $\mathbf{M} \in \mathcal{B}(N, r)$, as follows.

Proposition 4.2. *Let $\mathbf{M} \in \mathcal{B}(N, r)$. Then, there exists $\mathbf{U} \in \mathbb{R}^{r' \times rN}$ for some $r \leq r' \leq rN$ such that $\mathbf{M} = \mathbf{U}^\top \mathbf{U}$, where*

$$\mathbf{U} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \cdots & \mathbf{S}_N \\ \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{R}_N \end{bmatrix} \quad (50)$$

for some $\mathbf{S}_i \in \mathbb{R}_{\text{sym}}^{r \times r}$, $\mathbf{S}_1 = \mathbf{I}_r$, $\mathbf{R}_i \in \mathbb{R}^{(r'-r) \times r}$, $\mathbf{R}_1 = \mathbf{0}$, which satisfy the relations

$$\mathbf{S}_i^2 + \mathbf{R}_i^\top \mathbf{R}_i = \mathbf{I}_r, \quad (51)$$

$$\mathbf{S}_i \mathbf{S}_j - \mathbf{S}_j \mathbf{S}_i + \mathbf{R}_i^\top \mathbf{R}_j - \mathbf{R}_j^\top \mathbf{R}_i = \mathbf{0}. \quad (52)$$

(The latter relations encode the conditions $\mathbf{M}_{[ii]} = \mathbf{I}_r$ and $\mathbf{M}_{[ij]}^\top = \mathbf{M}_{[ij]}$, respectively.) Combining Proposition 4.1's Claim 3 and Proposition 4.2, we find that

$$\mathbf{v}_i = \mathbf{M}_{[i1]} \mathbf{v}_1 = \mathbf{S}_i \mathbf{v}_1, \quad (53)$$

$$\mathbf{v}_1 = \mathbf{M}_{[1i]} \mathbf{v}_i = \mathbf{S}_i \mathbf{v}_i. \quad (54)$$

We may therefore expand the entries of \mathbf{Y} in terms of the matrices \mathbf{S}_i and \mathbf{R}_i and the vector \mathbf{v}_1 :

$$\begin{aligned} Y_{(ij)(k\ell)} &= \mathbf{v}_i^\top \mathbf{M}_{[jk]} \mathbf{v}_\ell \\ &= \mathbf{v}_1^\top \mathbf{S}_i (\mathbf{S}_j \mathbf{S}_k + \mathbf{R}_j^\top \mathbf{R}_k) \mathbf{S}_\ell \mathbf{v}_1 \\ &= \mathbf{v}_1^\top \mathbf{S}_i \mathbf{S}_j \mathbf{S}_k \mathbf{S}_\ell \mathbf{v}_1 + \mathbf{v}_1^\top \mathbf{S}_i \mathbf{R}_j^\top \mathbf{R}_k \mathbf{S}_\ell \mathbf{v}_1. \end{aligned} \quad (55)$$

To show the first two equalities of (48), it then suffices to show that for any $i, j \in [N]$, we have

$$\mathbf{S}_i \mathbf{S}_j \mathbf{v}_1 \stackrel{?}{=} \mathbf{S}_j \mathbf{S}_i \mathbf{v}_1, \quad (56)$$

$$\mathbf{R}_i \mathbf{S}_j \mathbf{v}_1 \stackrel{?}{=} \mathbf{R}_j \mathbf{S}_i \mathbf{v}_1. \quad (57)$$

Observe first that, by (53) and (54), we have

$$\mathbf{S}_i^2 \mathbf{v}_1 = \mathbf{v}_1. \quad (58)$$

Taking (51) as a quadratic form with \mathbf{v}_1 , we find

$$1 = \|\mathbf{v}_1\|_2^2 = \mathbf{v}_1^\top \mathbf{S}_i^2 \mathbf{v}_1 + \|\mathbf{R}_i \mathbf{v}_1\|_2^2 = 1 + \|\mathbf{R}_i \mathbf{v}_1\|_2^2, \quad (59)$$

hence $\mathbf{R}_i \mathbf{v}_1 = \mathbf{0}$ for all $i \in [N]$. Then, multiplying (52) on the right by \mathbf{v}_1 establishes (56).

Next, taking (51) as a quadratic form with $\mathbf{v}_i = \mathbf{S}_i \mathbf{v}_1$, we find

$$1 = \|\mathbf{v}_i\|_2^2 = \|\mathbf{S}_i \mathbf{v}_1\|_2^2 + \|\mathbf{R}_i \mathbf{v}_i\|_2^2 = 1 + \|\mathbf{R}_i \mathbf{v}_i\|_2^2, \quad (60)$$

so $\mathbf{R}_i \mathbf{S}_i \mathbf{v}_1 = \mathbf{R}_i \mathbf{v}_i = \mathbf{0}$ for each $i \in [N]$ as well. Also, evaluating (51) as a quadratic form with $\mathbf{v}_j = \mathbf{S}_j \mathbf{v}_1$, we have

$$1 = \|\mathbf{v}_j\|_2^2 = \|\mathbf{S}_i \mathbf{S}_j \mathbf{v}_1\|_2^2 + \|\mathbf{R}_i \mathbf{S}_j \mathbf{v}_1\|_2^2. \quad (61)$$

Taking (51) as a bilinear form with $\mathbf{S}_i \mathbf{v}_1$ and $\mathbf{S}_j \mathbf{v}_1$ and using the two preceding observations gives

$$\begin{aligned} 0 &= \mathbf{v}_1^\top \mathbf{S}_j (\mathbf{S}_i \mathbf{S}_j - \mathbf{S}_j \mathbf{S}_i + \mathbf{R}_i^\top \mathbf{R}_j - \mathbf{R}_j^\top \mathbf{R}_i) \mathbf{S}_i \mathbf{v}_1 \\ &= \|\mathbf{S}_i \mathbf{S}_j \mathbf{v}_1\|_2^2 - 1 + \langle \mathbf{R}_i \mathbf{S}_j \mathbf{v}_1, \mathbf{R}_j \mathbf{S}_i \mathbf{v}_1 \rangle \\ &= -\|\mathbf{R}_i \mathbf{S}_j \mathbf{v}_1\|_2^2 + \langle \mathbf{R}_i \mathbf{S}_j \mathbf{v}_1, \mathbf{R}_j \mathbf{S}_i \mathbf{v}_1 \rangle. \end{aligned} \quad (62)$$

The same holds with indices i and j exchanged, so we find

$$\langle \mathbf{R}_i \mathbf{S}_j \mathbf{v}_1, \mathbf{R}_j \mathbf{S}_i \mathbf{v}_1 \rangle = \|\mathbf{R}_i \mathbf{S}_j \mathbf{v}_1\|_2^2 = \|\mathbf{R}_j \mathbf{S}_i \mathbf{v}_1\|_2^2 = \|\mathbf{R}_i \mathbf{S}_j \mathbf{v}_1\|_2 \|\mathbf{R}_j \mathbf{S}_i \mathbf{v}_1\|_2. \quad (63)$$

Thus the Cauchy-Schwarz inequality holds tightly between the vectors $\mathbf{R}_i \mathbf{S}_j \mathbf{v}_1$ and $\mathbf{R}_j \mathbf{S}_i \mathbf{v}_1$, so $\mathbf{R}_i \mathbf{S}_j \mathbf{v}_1 = \mathbf{R}_j \mathbf{S}_i \mathbf{v}_1$, establishing (57) and completing the proof.

4.3 Interpreting Theorem 2.15 as a Relaxation

Since \mathcal{E}_4^N may be seen as a relaxation of the cut polytope \mathcal{C}^N , one expects that the description of \mathcal{E}_4^N in terms of an SDP over the matrices of $\mathcal{B}(N, r)$, as stated in Theorem 2.15, should itself relax a description of \mathcal{C}^N in terms of a similar SDP with additional non-convex constraints. In this section, we show that the most naive such description one might expect is in fact incorrect, and give the correct description, which highlights an interesting connection with ideas from quantum information theory. The proofs of the results we give are deferred to Appendix C.

Naively, by analogy with the fact that if $\mathbf{X} \in \mathcal{E}_2^N$ with $\text{rank}(\mathbf{X}) = 1$ then $\mathbf{X} = \mathbf{x}\mathbf{x}^\top$ for $\mathbf{x} \in \{\pm 1\}^N$, one might expect that constraining the rank of $\mathbf{M} \in \mathcal{B}(N, r)$ in Theorem 2.15 to be as small as possible, namely to equal r , would give a description of \mathcal{C}^N . Unfortunately, as the following result shows, this only holds in one direction: if the Gram vector witness \mathbf{M} has rank r then the associated $\mathbf{X} \in \mathcal{C}^N$, but there exist $\mathbf{X} \in \mathcal{C}^N$ with $\text{rank}(\mathbf{X}) = r$ whose membership in \mathcal{E}_4^N does not admit a Gram vector witness \mathbf{M} with $\text{rank}(\mathbf{M}) = r$.

Proposition 4.3. *Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$, let $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$, and let $\mathbf{v} \in \mathbb{R}^{rN}$ be the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N$. Then, if $\sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N$ and there exists $\mathbf{M} \in \mathcal{B}(N, r)$ with $\text{rank}(\mathbf{M}) = r$ and $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$, then $\mathbf{X} \in \mathcal{C}^N$. On the other hand, if $N \notin \{1, 2\}$ and N is not divisible by 4, then $\mathbf{I}_N \in \mathcal{C}^N$ with $\mathbf{I}_N = \text{Gram}(\mathbf{e}_1, \dots, \mathbf{e}_N)$, but letting \mathbf{v} be the concatenation of $\mathbf{e}_1, \dots, \mathbf{e}_N$, there does not exist $\mathbf{M} \in \mathcal{B}(N, N)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$ and $\text{rank}(\mathbf{M}) = N$.*

(The unusual condition on the negative result that N be odd is probably superfluous if one searches for counterexamples other than the identity; the question is related to the relationship between the rank of a matrix in \mathcal{C}^N and the minimum number of cut matrices to whose convex hull it belongs, which, as is discussed in the proof, is known to be subtle.)

The correct way to “repair” this first attempt is quite surprising: the key condition on the Gram vector witness $\mathbf{M} \in \mathcal{B}(N, r)$ that is equivalent to $\mathbf{X} \in \mathcal{C}^N$ is not minimal rank, but *separability*, a notion studied mainly in quantum information theory. Related ideas will play a role in our derivation of constraints on the Gram vector witness \mathbf{M} defined in Theorem 2.15, but the full extent of this connection is still unclear and is an intriguing subject for future work.

Definition 4.4. A matrix $\mathbf{M} \in \mathbb{R}^{rN \times rN}$ with $\text{Tr}(\mathbf{M}) = 1$ is separable if there exist $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^N$ with $\|\mathbf{a}_i\|_2 = 1$, $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{R}^r$ with $\|\mathbf{b}_i\|_2 = 1$, and $\rho_1, \dots, \rho_m \geq 0$ with $\sum_i \rho_i = 1$ such that

$$\mathbf{M} = \sum_{i=1}^m \rho_i (\mathbf{a}_i \otimes \mathbf{b}_i)(\mathbf{a}_i \otimes \mathbf{b}_i)^\top. \quad (64)$$

If it is not possible to write \mathbf{M} in this way, \mathbf{M} is entangled. (More properly, \mathbf{M} is the density matrix representing, with respect to a particular choice of basis, a bipartite quantum state, and it is the state that is entangled or separable.) We write $\mathcal{B}_{\text{sep}}(N, r) \subseteq \mathcal{B}(N, r)$ for the matrices $\mathbf{M} \in \mathcal{B}(N, r)$ such that $\frac{1}{rN}\mathbf{M}$ is separable.

Proposition 4.5. Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$, let $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$, and let $\mathbf{v} \in \mathbb{R}^{rN}$ be the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N$. Then, $\mathbf{X} \in \mathcal{C}^N$ if and only if $\sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N$ and there exists $\mathbf{M} \in \mathcal{B}_{\text{sep}}(N, r)$ such that $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$.

By corollary, if $\mathbf{X} \in \mathcal{E}_4^N \setminus \mathcal{C}^N$, then any Gram vector witness \mathbf{M} (suitably scaled) must be the density matrix of an entangled state which, by the definition of $\mathcal{B}(N, r)$, has the *positive partial transpose (PPT)* property that its partial transpose remains psd. If the partial transpose of a density matrix of a state fails to be psd, it follows that the state is entangled, but the converse does not hold in general [60, 37]. The structure of states for which this test does not prove entanglement but which are nonetheless entangled has received considerable attention in the quantum information literature (see e.g. [52, 68, 50, 17], as well as [38, 10, 3] for more general discussion). It is therefore striking that these objects are, per our results, rather commonplace in SOS optimization—for every hypercube optimization problem for which degree 4 SOS is not tight (i.e. for which the optimizer $\mathbf{X}^* \in \mathcal{E}_4^N \setminus \mathcal{C}^N$), there is an underlying entangled PPT state that may be recovered from \mathbf{X}^* .

5 Constraints on Witnesses: Lemma 2.16 and Theorem 2.17

5.1 Proof of Lemma 2.16

Suppose that $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathcal{E}_4^N$ for some $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{S}^{r-1}$, \mathbf{v} is the concatenation of the \mathbf{v}_i , and $\mathbf{v}^\top \mathbf{M}^* \mathbf{v} = N^2$ for some $\mathbf{M}^* \in \mathcal{B}(N, r)$. Then, \mathbf{M}^* is an optimizer for the following SDP, described by Theorem 2.15:

$$\text{GramSDP}(\mathbf{v}_1, \dots, \mathbf{v}_N) := \left\{ \begin{array}{l} \text{maximize} \quad \langle \mathbf{v} \mathbf{v}^\top, \mathbf{M} \rangle \\ \text{subject to} \quad \mathbf{M} \succeq 0, \\ \quad \quad \quad \mathbf{M}_{[ii]} = \mathbf{I}_r, \\ \quad \quad \quad \mathbf{M}_{[ij]} = \mathbf{M}_{[ij]}^\top \text{ for } i \neq j. \end{array} \right\}. \quad (65)$$

We next apply basic convex optimization results to this SDP. Background on these general facts may be found in [8, 12]. First, we obtain the dual SDP

$$\text{GramSDP}^*(\mathbf{v}_1, \dots, \mathbf{v}_N) := \left\{ \begin{array}{l} \text{minimize} \quad \text{Tr}(\mathbf{D}) \\ \text{subject to} \quad \mathbf{D} \succeq \mathbf{v} \mathbf{v}^\top, \\ \quad \quad \quad \mathbf{D}_{[ij]} = -\mathbf{D}_{[ij]}^\top \text{ for } i \neq j. \end{array} \right\}. \quad (66)$$

It is simple to verify that the Slater condition holds, implying strong duality between the SDPs (65) and (66), whereby $\text{GramSDP}(\mathbf{v}_1, \dots, \mathbf{v}_N) = \text{GramSDP}^*(\mathbf{v}_1, \dots, \mathbf{v}_N) = N^2$. If \mathbf{M}^* and \mathbf{D}^* are primal and dual variables achieving the optimal values of GramSDP and GramSDP* respectively, then *complementary slackness* must hold between them, $\mathbf{M}^*(\mathbf{D}^* - \mathbf{v} \mathbf{v}^\top) = \mathbf{0}$.

The key to Lemma 2.16 is that, while constructing \mathbf{M}^* achieving a value of N^2 in GramSDP from $\mathbf{v}_1, \dots, \mathbf{v}_N$ (when their Gram matrix belongs to \mathcal{E}_4^N) is difficult (indeed, by the more detailed part of Theorem 2.15 it is equivalent to constructing the degree 4 pseudomoments themselves), constructing \mathbf{D}^* achieving a value of N^2 in GramSDP* turns out to be straightforward.

The construction uses the *partial transpose* operation, $\mathbf{A} \mapsto \mathbf{A}^\Gamma$ mapping $\mathbb{R}^{rN \times rN} \rightarrow \mathbb{R}^{rN \times rN}$, which transposes every $r \times r$ block of an $rN \times rN$ matrix:

$$\mathbf{A}^\Gamma := \left[\mathbf{A}_{[ij]}^\top \right]_{i,j=1}^N \quad \text{for } \mathbf{A} \in \mathbb{R}^{rN \times rN}. \quad (67)$$

We then define \mathbf{D}^* as

$$\mathbf{D}^* := \mathbf{v}\mathbf{v}^\top - (\mathbf{v}\mathbf{v}^\top)^\Gamma + \mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top). \quad (68)$$

We have $\text{Tr}(\mathbf{D}^*) = \text{Tr}(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top)) = N^2$, and for $i \neq j$, $\mathbf{D}_{[ij]} = \mathbf{v}_i\mathbf{v}_j^\top - \mathbf{v}_j\mathbf{v}_i^\top$, which is antisymmetric as required. The final feasibility condition $\mathbf{D}^* \succeq \mathbf{v}\mathbf{v}^\top$ is equivalent to $(\mathbf{v}\mathbf{v}^\top)^\Gamma \preceq \mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top)$, which is simple to check directly: let $\mathbf{x} \in \mathbb{R}^{rN}$ be the concatenation of $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}^r$, then

$$\begin{aligned} \mathbf{x}^\top (\mathbf{v}\mathbf{v}^\top)^\Gamma \mathbf{x} &= \sum_{i=1}^N \sum_{j=1}^N \langle \mathbf{x}_i, \mathbf{v}_j \rangle \langle \mathbf{x}_j, \mathbf{v}_i \rangle \\ &\leq \sum_{i=1}^N \sum_{j=1}^N \langle \mathbf{x}_i, \mathbf{v}_j \rangle^2 \\ &= \sum_{i=1}^N \mathbf{x}_i^\top (\mathbf{V}\mathbf{V}^\top) \mathbf{x}_i \\ &= \mathbf{x}^\top (\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top)) \mathbf{x}. \end{aligned}$$

Thus, by complementary slackness an optimizer \mathbf{M}^* in GramSDP must have positive eigenvectors in $\ker(\mathbf{D}^* - \mathbf{v}\mathbf{v}^\top) = \ker(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma)$.

Remarkably, this subspace may be calculated exactly to produce the result of Lemma 2.16. This calculation hinges on some elementary but perhaps not widely known facts of linear algebra that originate in applications to quantum information theory. The first is the following, a rewriting of the singular value decomposition.

Proposition 5.1 (Schmidt Decomposition, Section 2.2.2 of [3]). *Let $r \leq N$, $\mathbf{V} \in \mathbb{R}^{r \times N}$ having singular value decomposition $\mathbf{V} = \sum_{i=1}^r \sigma_i \mathbf{y}_i \mathbf{z}_i^\top$, where the $\mathbf{y}_i \in \mathbb{R}^r$ and $\mathbf{z}_i \in \mathbb{R}^N$ each form orthonormal sets and $\sigma_i \geq 0$. Then,*

$$\text{vec}(\mathbf{V}) = \sum_{i=1}^r \sigma_i \mathbf{z}_i \otimes \mathbf{y}_i. \quad (69)$$

Note that in our case, $\mathbf{v} = \text{vec}(\mathbf{V})$.

This representation makes it convenient to work with the partial transpose; in particular, using the Schmidt decomposition, it is possible to diagonalize the partial transpose of a rank one matrix explicitly, as follows. (This result appears to be folkloric in the quantum information literature; the references we give are unlikely to be the earliest.)

Proposition 5.2 (Lemma III.3 of [34]; Lemma 1 of [39]). *Let $\mathbf{V} \in \mathbb{R}^{r \times N}$ with $r \leq N$ and $\mathbf{V} = \sum_{i=1}^r \sigma_i \mathbf{y}_i \mathbf{z}_i^\top$ where the $\mathbf{y}_i \in \mathbb{R}^r$ and $\mathbf{z}_i \in \mathbb{R}^N$ each form orthonormal sets and $\sigma_i \geq 0$. Let*

$\mathbf{v} = \text{vec}(\mathbf{V})$. Then,

$$(\mathbf{v}\mathbf{v}^\top)^\Gamma = \sum_{i=1}^r \sigma_i^2 \mathbf{d}_i \mathbf{d}_i^\top + \sum_{1 \leq i < j \leq r} \sigma_i \sigma_j \mathbf{s}_{ij} \mathbf{s}_{ij}^\top - \sum_{1 \leq i < j \leq r} \sigma_i \sigma_j \mathbf{a}_{ij} \mathbf{a}_{ij}^\top \quad (70)$$

where

$$\mathbf{d}_i = \mathbf{z}_i \otimes \mathbf{y}_i, \quad (71)$$

$$\mathbf{s}_{ij} = \frac{1}{\sqrt{2}} (\mathbf{z}_i \otimes \mathbf{y}_j + \mathbf{z}_j \otimes \mathbf{y}_i), \quad (72)$$

$$\mathbf{a}_{ij} = \frac{1}{\sqrt{2}} (\mathbf{z}_i \otimes \mathbf{y}_j - \mathbf{z}_j \otimes \mathbf{y}_i). \quad (73)$$

The r^2 vectors $\mathbf{d}_i, \mathbf{s}_{ij}, \mathbf{a}_{ij}$ moreover have unit norm and are mutually orthogonal, so (70) is a spectral decomposition (up to the removal of any terms whose coefficient is zero if \mathbf{V} is not full rank).

With this, it is straightforward to compute the subspace we are interested in, since $\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top)$ can also be diagonalized explicitly in a basis of the same Kronecker products $\mathbf{z}_i \otimes \mathbf{y}_j$.

Proposition 5.3. *Let $\mathbf{V} \in \mathbb{R}^{r \times N}$ with $r \leq N$ have full rank, and let $\mathbf{v} = \text{vec}(\mathbf{V})$. Then,*

$$\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) \succeq (\mathbf{v}\mathbf{v}^\top)^\Gamma. \quad (74)$$

The subspace on which this inequality is tight is given by

$$\ker \left(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma \right) = \{ \text{vec}(\mathbf{S}\mathbf{V}) : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r} \} =: V_{\text{sym}}. \quad (75)$$

Letting $\mathbf{V} = \sum_{i=1}^r \sigma_i \mathbf{y}_i \mathbf{z}_i^\top$ for $\mathbf{y}_i \in \mathbb{R}^r$ an orthonormal basis, $\mathbf{z}_i \in \mathbb{R}^N$ an orthonormal set, and $\sigma_i > 0$ be the singular decomposition, an orthonormal basis for V_{sym} is given by the $\frac{r(r+1)}{2}$ vectors

$$\mathbf{z}_i \otimes \mathbf{y}_i \text{ for } 1 \leq i \leq r, \quad (76)$$

$$\frac{1}{\sqrt{\sigma_i^2 + \sigma_j^2}} (\sigma_i \mathbf{z}_i \otimes \mathbf{y}_j + \sigma_j \mathbf{z}_j \otimes \mathbf{y}_i) \text{ for } 1 \leq i < j \leq r. \quad (77)$$

We provide proofs of the preceding three Propositions in Appendix D, giving more details on Proposition 5.3 since it is the only one of these results that appears to be original.

From the previous discussion of complementary slackness and Proposition 5.3, we find that all eigenvectors with positive eigenvalue of \mathbf{M} must belong to V_{sym} . To obtain from this the statement of Lemma 2.16, first note that if $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$ then by Proposition 4.1 $\mathbf{M} \mathbf{v} = N \mathbf{v}$, so $\mathbf{M} = \mathbf{v}\mathbf{v}^\top + \mathbf{M}'$ for some $\mathbf{M}' \succeq \mathbf{0}$. Suppose that $\mathbf{w} \in \mathbb{R}^{rN}$ is an eigenvector of \mathbf{M}' with eigenvalue $\lambda > 0$. Then, $\mathbf{w} \in V_{\text{sym}}$ by the above reasoning, so $\mathbf{w} = \text{vec}(\mathbf{S}\mathbf{V})$ for some $\mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}$. Also,

$$\mathbf{I}_r = \mathbf{M}_{[ii]} \succeq (\mathbf{v}\mathbf{v}^\top + \lambda \mathbf{w}\mathbf{w}^\top)_{[ii]} = \mathbf{v}_i \mathbf{v}_i^\top + \lambda \mathbf{S} \mathbf{v}_i \mathbf{v}_i^\top \mathbf{S}, \quad (78)$$

and taking this as a quadratic form with \mathbf{v}_i shows that $\mathbf{v}_i \mathbf{S} \mathbf{v}_i = 0$. Since this holds for each $i \in [r]$, we obtain the conclusion of Lemma 2.16, that

$$\mathbf{w} \in V'_{\text{sym}} := \left\{ \text{vec}(\mathbf{S}\mathbf{V}) : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}, \mathbf{v}_i^\top \mathbf{S} \mathbf{v}_i = 0 \text{ for } i \in [r] \right\}. \quad (79)$$

5.2 Proof of Theorem 2.17

Suppose $\mathbf{X} \in \mathcal{E}_4^N$ with $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$, and $\mathbf{v}_i \in \mathbb{R}^r$ with $r = \text{rank}(\mathbf{X})$. Then if $\mathbf{V} \in \mathbb{R}^{r \times N}$ has the \mathbf{v}_i as its columns, \mathbf{V} is full-rank. If $\mathbf{Y} \in \mathbb{R}^{N^2 \times N^2}$ is any degree 4 pseudomoment matrix extending \mathbf{X} , then there is $\mathbf{M} \in \mathcal{B}(N, r)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$. Suppose $r' = \text{rank}(\mathbf{M})$, then let us write the spectral decomposition $\mathbf{M} = \mathbf{v} \mathbf{v}^\top + \sum_{m=1}^{r'-1} \lambda_m \mathbf{w}_m \mathbf{w}_m^\top$ for some $\lambda_m > 0$.

By Lemma 2.16, $\mathbf{w}_m \in V'_{\text{sym}}$. Therefore, $\mathbf{w}_m = \text{vec}(\mathbf{S}_m \mathbf{V})$ for some $\mathbf{S}_m \in \mathbb{R}_{\text{sym}}^{r \times r}$ with $\mathbf{v}_i^\top \mathbf{S}_m \mathbf{v}_i = 0$ for all $i \in [N], m \in [r' - 1]$. By (13) from Theorem 2.15, we may therefore expand

$$\mathbf{Y} = \tilde{\mathbf{v}} \tilde{\mathbf{v}}^\top + \sum_{m=1}^{r'-1} \lambda_m \tilde{\mathbf{w}}_m \tilde{\mathbf{w}}_m^\top, \quad (80)$$

$$(\tilde{\mathbf{v}})_{(ij)} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle, \quad (81)$$

$$(\tilde{\mathbf{w}}_m)_{(ij)} = \langle \mathbf{S}_m \mathbf{v}_i, \mathbf{v}_j \rangle. \quad (82)$$

Thus, we simply have $\tilde{\mathbf{v}} = \text{vec}(\mathbf{X})$ and $\tilde{\mathbf{w}}_m = \text{vec}(\mathbf{V}^\top \mathbf{S}_m \mathbf{V})$.

The statement of Theorem 2.17 comes from combining this with the following previous result about the facial geometry of \mathcal{E}_2^N .

Proposition 5.4 (Theorem 1(a) of [53]). *Let $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathcal{E}_2^N$ for $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{S}^{r-1}$ having $\text{rank}(\mathbf{X}) = r$, and let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns, so that $\mathbf{X} = \mathbf{V}^\top \mathbf{V}$. Then,*

$$\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}) = \left\{ \mathbf{V}^\top \mathbf{S} \mathbf{V} : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r} \right\} \cap \left\{ \mathbf{A} \in \mathbb{R}^{N \times N} : \text{diag}(\mathbf{A}) = \mathbf{0} \right\} \quad (83)$$

$$= \left\{ \mathbf{V}^\top \mathbf{S} \mathbf{V} : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}, \mathbf{v}_i^\top \mathbf{S} \mathbf{v}_i = 0 \text{ for } i \in [N] \right\}. \quad (84)$$

Therefore, continuing the reasoning above, we find that for each $m \in [r' - 1]$, $\tilde{\mathbf{w}}_m = \text{vec}(\mathbf{A}_m)$ for some $\mathbf{A}_m \in \text{pert}_{\mathcal{E}_2^N}(\mathbf{X})$. Hence, every eigenvector of $\mathbf{Y} - \text{vec}(\mathbf{X}) \text{vec}(\mathbf{X})^\top$ having nonzero eigenvalue must lie in $\text{vec}(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}))$, establishing the first part of the result.

The second part of the result controls $\text{rank}(\mathbf{Y}) \leq r'$. By the first part of the result,

$$r' \leq \dim \left(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}) \right) + 1, \quad (85)$$

so it suffices to compute the right-hand side. Since \mathbf{V} is full-rank, the map $\mathbf{S} \mapsto \mathbf{V}^\top \mathbf{S} \mathbf{V}$ is injective, so this may be computed as

$$\dim \left(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}) \right) = \dim \left(\text{span} \left(\left\{ \mathbf{v}_i \mathbf{v}_i^\top \right\}_{i=1}^N \right)^\perp \right) = \frac{r(r+1)}{2} - \dim \left(\text{span} \left(\left\{ \mathbf{v}_i \mathbf{v}_i^\top \right\}_{i=1}^N \right) \right). \quad (86)$$

Since $\text{Gram}(\mathbf{v}_1 \mathbf{v}_1^\top, \dots, \mathbf{v}_N \mathbf{v}_N^\top) = \mathbf{X}^{\odot 2}$, we equivalently have

$$\dim \left(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}) \right) = \frac{r(r+1)}{2} - \text{rank}(\mathbf{X}^{\odot 2}), \quad (87)$$

a previously known corollary of Proposition 5.4 used in [53, 49].

The final part of the result concerns the special case where $\mathbf{X} \in \mathcal{E}_2^N$ is an extreme point, whereby $\dim(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X})) = 0$. Then, if \mathbf{Y} is a degree 4 pseudomoment matrix extending \mathbf{X} we have $\text{rank}(\mathbf{Y}) = r' = 1$, so $\text{rank}(\mathbf{X}) = 1$ as well since \mathbf{X} is a minor of \mathbf{Y} . Since $\mathbf{X} \in \mathcal{E}_2^N$, in fact $\mathbf{X} = \mathbf{x} \mathbf{x}^\top$ for some $\mathbf{x} \in \{\pm 1\}^N$, and it is simple to check that the only possible degree 4 extension of rank one is then $\mathbf{Y} = (\mathbf{x} \otimes \mathbf{x})(\mathbf{x} \otimes \mathbf{x})^\top$.

6 Examples from Equiangular Tight Frames: Theorem 2.19

Before giving the proofs of our results on ETFs, we first point out a general convenience of working with the Gram matrices of UNTFs through Theorem 2.15. Suppose $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$ and the $\mathbf{v}_i \in \mathbb{R}^r$ form a UNTF. Let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns. Suppose also that $\mathbf{M} \in \mathcal{B}(N, r)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$. Lemma 2.16 then ensures that the eigenvectors of \mathbf{M} with positive eigenvalue lie in the subspace of vectors of the form $\text{vec}(\mathbf{S}\mathbf{V})$ for $\mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}$. In the case where $\mathbf{v}_1, \dots, \mathbf{v}_N$ form a UNTF, we show that in fact this mapping is, up to scaling, an isometry.

Definition 6.1. For $\mathbf{V} \in \mathbb{R}^{r \times N}$, let us write $\mathcal{V}_{\mathbf{V}} : \mathbb{R}_{\text{sym}}^{r \times r} \rightarrow \mathbb{R}^{r \times N}$ for the map $\mathcal{V}_{\mathbf{V}}(\mathbf{S}) = \sqrt{\frac{r}{N}} \text{vec}(\mathbf{S}\mathbf{V})$. When the matrix \mathbf{V} is clear from context, we will drop the subscript \mathbf{V} .

Proposition 6.2. Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{S}^{r-1}$ form a UNTF and let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns. Then, the mapping $\mathcal{V} = \mathcal{V}_{\mathbf{V}}$ is a linear isometry between $\mathbb{R}_{\text{sym}}^{r \times r}$ and $\{\text{vec}(\mathbf{S}\mathbf{V}) : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}\} \subset \mathbb{R}^{r \times N}$, if $\mathbb{R}_{\text{sym}}^{r \times r}$ is endowed with the Frobenius inner product $\langle \mathbf{S}, \mathbf{S}' \rangle = \text{Tr}(\mathbf{S}\mathbf{S}')$.

Proof. To check that inner products are preserved, we compute:

$$\begin{aligned} \langle \mathcal{V}(\mathbf{S}), \mathcal{V}(\mathbf{S}') \rangle &= \frac{r}{N} \sum_{i=1}^N \langle \mathbf{S}\mathbf{v}_i, \mathbf{S}'\mathbf{v}_i \rangle \\ &= \frac{r}{N} \sum_{i=1}^N \mathbf{v}_i^\top \mathbf{S}\mathbf{S}'\mathbf{v}_i \\ &= \text{Tr} \left(\mathbf{S}\mathbf{S}' \left(\frac{r}{N} \mathbf{V}\mathbf{V}^\top \right) \right) \\ &= \langle \mathbf{S}, \mathbf{S}' \rangle. \end{aligned} \tag{88}$$

Clearly \mathcal{V} is linear, injectivity follows from the \mathbf{v}_i forming a spanning set, and surjectivity follows from the definition of the target space. \square

Similarly, Theorem 2.15 shows that \mathbf{Y} can be produced from \mathbf{M} by conjugating as

$$\mathbf{Y} = (\mathbf{I}_N \otimes \mathbf{V})^\top \mathbf{M} (\mathbf{I}_N \otimes \mathbf{V}) = \frac{N}{r} \left(\mathbf{I}_N \otimes \sqrt{\frac{r}{N}} \mathbf{V} \right)^\top \mathbf{M} \left(\mathbf{I}_N \otimes \sqrt{\frac{r}{N}} \mathbf{V} \right), \tag{89}$$

where in the latter expression the matrix $\mathbf{I}_N \otimes \sqrt{\frac{r}{N}} \mathbf{V}$ has orthonormal rows, so \mathbf{Y} is also merely a scaled and rotated copy of \mathbf{M} , embedded in a higher-dimensional space. In particular, the spectrum of \mathbf{Y} is the spectrum of \mathbf{M} , scaled up by $\frac{N}{r}$.

6.1 Proof of Theorem 2.19

In this section we will prove the necessary and sufficient condition for the Gram matrix of an ETF to lie in \mathcal{E}_4^N . Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$ form an ETF, let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns, let $\mathbf{v} = \text{vec}(\mathbf{V})$ be the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N$, and let $\mathbf{X} = \mathbf{V}^\top \mathbf{V} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$. Then, our result is that $\mathbf{X} \in \mathcal{E}_4^N$ if and only if $N < \frac{r(r+1)}{2}$ or $r = 1$. If $r = 1$, then each \mathbf{v}_i is a scalar equal to ± 1 , so $\mathbf{X} \in \mathcal{C}^N$. Thus, it suffices to restrict our attention to $r > 1$.

First, we recall a classical result on equiangular lines (not necessarily forming a tight frame) which shows that this result only excludes one extremal case. We also include its elegant proof, since similar ideas will be involved in our argument.

Proposition 6.3 (Gerzon Bound [51]). *If $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{S}^{r-1}$ and $|\langle \mathbf{v}_i, \mathbf{v}_j \rangle| = \alpha < 1$ for all $i, j \in [N]$ with $i \neq j$, then $N \leq \frac{r(r+1)}{2}$.*

Proof. For all $i \neq j$, $\langle \mathbf{v}_i \mathbf{v}_i^\top, \mathbf{v}_j \mathbf{v}_j^\top \rangle = \alpha^2$. Thus,

$$\text{Gram}(\mathbf{v}_1 \mathbf{v}_1^\top, \dots, \mathbf{v}_N \mathbf{v}_N^\top) = (1 - \alpha^2) \mathbf{I}_N + \alpha^2 \mathbf{1} \mathbf{1}^\top, \quad (90)$$

which is non-singular. The $\mathbf{v}_i \mathbf{v}_i^\top$ are then linearly independent, so $N \leq \dim(\mathbb{R}_{\text{sym}}^{r \times r}) = \frac{r(r+1)}{2}$. \square

By Theorem 2.17, the negative direction of Theorem 2.19 immediately follows: if $N = \frac{r(r+1)}{2}$, then the $\mathbf{v}_i \mathbf{v}_i^\top$ span $\mathbb{R}_{\text{sym}}^{r \times r}$, so by Proposition 5.4 \mathbf{X} is an extreme point of \mathcal{E}_2^N , thus \mathbf{X} cannot belong to \mathcal{E}_4^N unless $\text{rank}(\mathbf{X}) = 1$, which is a contradiction if $r > 1$.

The positive direction with $r > 1$ is the more difficult part of the result. We proceed by explicitly constructing $\mathbf{M} \in \mathcal{B}(N, r)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$. The construction is optimistic: we consider the simplest possible choice for \mathbf{M} respecting the constraint of Lemma 2.16. The Lemma forces $\mathbf{M} = \mathbf{v} \mathbf{v}^\top + \mathbf{M}'$ where $\mathbf{M}' \succeq 0$ with all of its eigenvectors with positive eigenvalue lying in the subspace V'_{sym} . We then simply choose \mathbf{M}' to equal a constant multiple of $\mathbf{P}_{V'_{\text{sym}}}$. Choosing the constant factor such that $\text{Tr}(\mathbf{M}) = rN$, we obtain the candidate

$$\mathbf{M} := \mathbf{v} \mathbf{v}^\top + \frac{(r-1)N}{\frac{r(r+1)}{2} - N} \mathbf{P}_{V'_{\text{sym}}}. \quad (91)$$

If we could show that $\mathbf{M}_{[ii]} = \mathbf{I}_r$ and $\mathbf{M}_{[ij]}^\top = \mathbf{M}_{[ij]}$ for all $i, j \in [N]$, then the proof would be complete.

Surprisingly, the naive construction (91) does satisfy these properties. This may be verified by calculating $\mathbf{P}_{V'_{\text{sym}}}$ explicitly, a calculation we perform in detail in Appendix E but outline briefly here. Recall that

$$V'_{\text{sym}} := \left\{ \text{vec}(\mathbf{S} \mathbf{V}) : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}, \mathbf{v}_i^\top \mathbf{S} \mathbf{v}_i = 0 \text{ for } i \in [N] \right\} \subset \mathbb{R}^{rN}. \quad (92)$$

The basic idea is then to write $\mathbf{P}_{V'_{\text{sym}}} \mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}^{rN}$ as $\text{vec}(\mathbf{S} \mathbf{V})$ for \mathbf{S} solving the least-squares optimization problem for the orthogonal projection of \mathbf{y} . We then solve this optimization explicitly with Lagrange multipliers. Determining the Lagrange multipliers in turn amounts to inverting the matrix $\mathbf{X}^{\odot 2}$. Fortunately, for an ETF, as noted previously in the introduction and in the proof of Proposition 6.3, this matrix has a simple structure, making the calculation tractable. In this way, we obtain formulae for the blocks of $\mathbf{P}_{V'_{\text{sym}}}$ (see Corollary E.3), after which it is straightforward to check that $\mathbf{M} \in \mathcal{B}(N, r)$.

Finally, using the relation (14) between the blocks of \mathbf{M} and the degree 4 pseudomoments, we recover the elegant formula for the degree 4 pseudomoments:

$$Y_{(ij)(k\ell)} = \frac{\frac{r(r-1)}{2}}{\frac{r(r+1)}{2} - N} (X_{ij} X_{k\ell} + X_{ik} X_{j\ell} + X_{il} X_{jk}) - \frac{r^2 \left(1 - \frac{1}{N}\right)}{\frac{r(r+1)}{2} - N} \sum_{m=1}^N X_{im} X_{jm} X_{km} X_{\ell m}. \quad (93)$$

This derivation is a rather egregious instance of “bookkeeping for a miracle” [19], and it certainly remains an open question to provide an intuitive explanation for why any ETF Gram matrices ought to belong to \mathcal{E}_4^N at all, or for the structure of the remarkably symmetric formula (93). We remark additionally that, by the comments at the beginning of this section, the spectrum of \mathbf{Y} described by (93) is the same as the spectrum of \mathbf{M} with a constant scaling, and thus is also simple: \mathbf{Y} equals the rank one matrix $\text{vec}(\mathbf{X}) \text{vec}(\mathbf{X})^\top$ plus a constant multiple of the projection matrix onto the subspace $\text{vec}(\text{pert}_{\mathcal{E}_2^N}(\mathbf{X}))$.

7 Applications

7.1 Schläfli Inequalities: Theorem 2.20

In this section we will describe the computer-assisted verification of the inequalities (25) and some ancillary results. First, we review a connection between ETFs and strongly regular graphs (SRGs). (In fact, there are two distinct correspondences between ETFs and SRGs: the one we will use applies to arbitrary ETFs and is described in [28], while the other applies only to ETFs with a certain additional symmetry and is described in [26].)

Definition 7.1. A graph $G = (V, E)$ is a strongly regular graph with parameters (v, k, λ, μ) , abbreviated $\text{srg}(v, k, \lambda, \mu)$, if $|V| = v$, G is k -regular, every $x, y \in V$ that are adjacent have λ common neighbors, and every $x, y \in V$ that are not adjacent have μ common neighbors.

Proposition 7.2 (Theorem 3.1 of [28]). Let $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$ form an ETF with $N > r$, suppose that for all $i \in [N] \setminus \{1\}$ we have $\langle \mathbf{v}_1, \mathbf{v}_i \rangle > 0$, and let $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$. Define the graph G on vertices in $[N] \setminus \{1\}$ where i and j are adjacent if and only if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle > 0$. Then, G is an $\text{srg}(v, k, \lambda, \mu)$ with parameters

$$v = N - 1, \tag{94}$$

$$k = \frac{N}{2} - 1 + \left(\frac{N}{2r} - 1 \right) \sqrt{\frac{r(N-1)}{N-r}}, \tag{95}$$

$$\mu = \frac{k}{2}, \tag{96}$$

$$\lambda = \frac{3k - v - 1}{2}. \tag{97}$$

Note that the assumption that $\langle \mathbf{v}_1, \mathbf{v}_i \rangle > 0$ for all $i \neq 1$ is not a substantial restriction, since any vector in an ETF may be negated to produce another essentially equivalent ETF.

In our case, an ETF on 28 vectors in \mathbb{R}^7 corresponds to an $\text{srg}(27, 16, 10, 8)$. By the result of [66], this graph is unique, so we may take it by definition to be the Schläfli graph (a more natural geometric description is given in the previous reference). Consequently, since by negating some vectors every ETF can be put into the “canonical” form where $\langle \mathbf{v}_1, \mathbf{v}_i \rangle > 0$ for all $i \neq 1$, we obtain the following uniqueness result.

Proposition 7.3. Let $\mathbf{v}_1, \dots, \mathbf{v}_{28}$ and $\mathbf{w}_1, \dots, \mathbf{w}_{28}$ be two ETFs in \mathbb{R}^7 . Then, there exist signs $1 = s_1, s_2, \dots, s_{28} \in \{\pm 1\}$ and $\mathbf{Q} \in \mathcal{O}(7)$ such that $\mathbf{w}_i = s_i \mathbf{Q} \mathbf{v}_i$ for each $i \in [28]$.

Since if $\mathbf{X} \in \mathcal{E}_4^N$ then $\mathbf{D}\mathbf{X}\mathbf{D} \in \mathcal{E}_4^N$ for any $\mathbf{D} = \text{diag}(\mathbf{d})$ with $\mathbf{d} \in \{\pm 1\}^N$, it suffices to fix a single ETF of 28 vectors in \mathbb{R}^7 and check (25), and the result will follow for all ETFs of the same dimensions. Thus, let us fix $\mathbf{v}_1, \dots, \mathbf{v}_{28} \in \mathbb{R}^7$ forming an ETF with $\langle \mathbf{v}_1, \mathbf{v}_i \rangle > 0$ for all $i \neq 1$, and let $\mathbf{Z} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_{28})$. Let G be the graph on [28] where i and j are adjacent if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle > 0$, so that G is the Schläfli graph with one extra vertex added that is attached to every other vertex. We will write $G|_S$ for the subgraph induced by G on the set of vertices S .

We show (25) by producing a $\mathbf{0} \preceq \mathbf{A} \in \mathbb{R}^{N^2 \times N^2}$ such that for any \mathbf{Y} a degree 4 pseudomoment matrix extending some \mathbf{X} a degree 2 pseudomoment matrix,

$$0 \leq \langle \mathbf{A}, \mathbf{Y} \rangle = 112 - \sum_{1 \leq i < j \leq 28} \text{sgn}(Z_{ij}) X_{ij}. \tag{98}$$

The construction of \mathbf{A} is based on studying the results of numerical experiments. We identify the constants appearing in \mathbf{A} as

$$\gamma_1 := \frac{1}{126}, \quad (99)$$

$$\gamma_2 := \frac{1}{36}, \quad (100)$$

$$\kappa_1 := \frac{2}{9}, \quad (101)$$

$$\kappa_2 := \frac{1}{28}. \quad (102)$$

With this, we define

$$A_{(ij)(k\ell)} := \begin{cases} 0 & : |\{i, j, k, \ell\}| = 4, \\ -\text{sgn}(Z_{k\ell})\gamma_1 & : i = j, k \neq \ell, \\ \gamma_2 & : i = k, j \neq \ell, |E(G|_{\{i,j,\ell}\})| = 0, \\ \gamma_2 & : i = k, j \neq \ell, |E(G|_{\{i,j,\ell}\})| = 2, i \sim j, i \sim \ell, \\ -\gamma_2 & : i = k, j \neq \ell, |E(G|_{\{i,j,\ell}\})| = 2, j \sim \ell, \\ 0 & : i = k, j \neq \ell, |E(G|_{\{i,j,\ell}\})| \in \{1, 3\}, \\ -\text{sgn}(Z_{i\ell})\gamma_1 & : i = j = k, i \neq \ell, \\ \kappa_1 & : i = k, j = \ell, i \neq j, \\ \kappa_2 & : i = j, k = \ell. \end{cases} \quad (103)$$

$$A_{(i_1 i_2)(i_3 i_4)} = A_{(i_{\pi(1)} i_{\pi(2)})(i_{\pi(3)} i_{\pi(4)})} \text{ for } \mathbf{i} \in [N]^4, \pi \in \text{Sym}(4). \quad (104)$$

We then perform a computer verification that $\mathbf{A} \succeq \mathbf{0}$ using the `SageMath` software package for symbolic calculation of a Cholesky decomposition. Verifying that the equality of (98) holds is straightforward by counting the occurrences of various terms in $\langle \mathbf{A}, \mathbf{X} \rangle$. Accompanying code for reproducing the verification is available online.³ Of course, this proof technique is rather unsatisfying, and it is an open problem to provide a more principled description of \mathbf{A} and a conceptual proof of its positive semidefiniteness (both for this specific case and for the general case of maximal ETFs for any dimensions they may exist in).

7.2 Complexity of Parity Inequalities

In this section, we give the straightforward argument behind our proof of Corollary 2.22, a partial reproduction of the result of Laurent given in Proposition 2.21. The matrix $\mathbf{X}^{(N)}$ described there is the Gram matrix of the following type of ETF.

Definition 7.4. *The simplex ETF with parameter $N \geq 3$ is an ETF of N vectors in \mathbb{R}^r with $r = N - 1$, whose vectors point to the vertices of an equilateral simplex whose barycenter lies at the origin and whose vertices are unit distance from this barycenter. The coherence of the simplex ETF is $\alpha = \frac{1}{r}$, and the inner product of any two distinct vectors is $-\alpha$; that is, the Gram matrix $\mathbf{X}^{(N)}$ is*

$$\mathbf{X}^{(N)} = \left(1 + \frac{1}{N-1}\right) \mathbf{I}_N - \frac{1}{N-1} \mathbf{1}\mathbf{1}^\top. \quad (105)$$

When $N = 3$ and $r = 2$, then the simplex ETF is maximal, so $\mathbf{X}^{(3)} \notin \mathcal{E}_4^3$, but for all $N > 3$ we do have $\mathbf{X}^{(N)} \in \mathcal{E}_4^N$, and the extending degree 4 pseudomoment matrix may be computed directly

³See the second author's webpage at <http://www.kunisky.com/publications/deg-4-elliptope/>.

from (93) (especially simple in this case since the terms X_{ij} appearing in the summations only take on two different values), completing the proof of the Corollary.

We remark that, in our previous results on ETFs, the only technical calculation was that of the projection matrix $P_{V_{\text{sym}}}$, and the only particularly novel idea required was the optimistic construction of the Gram vector witness (91). In contrast, the original argument of [46] uses some rather powerful machinery from the general theory of association schemes and analytic identities for hypergeometric functions. We thus hope that our approach can be extended to view the higher-degree pseudomoment matrices used for the full result of Proposition 2.21 as Gram matrices as well, replacing these technicalities with simpler considerations of the geometry of the simplex ETFs.

Acknowledgements

We thank Jess Banks, Nicolas Boumal, Didier Henrion, Aida Khajavirad, Jean-Bernard Lasserre, Dustin Mixon, Cristopher Moore, and participants in the 2018 Princeton Day of Optimization for useful discussions.

References

- [1] Emmanuel Abbe, Afonso S Bandeira, Annina Bracher, and Amit Singer. Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Transactions on Network Science and Engineering*, 1(1):10–22, 2014.
- [2] Ahmed El Alaoui, Florent Krzakala, and Michael I Jordan. Fundamental limits of detection in the spiked wigner model. *arXiv preprint arXiv:1806.09588*, 2018.
- [3] Guillaume Aubrun and Stanisław J Szarek. *Alice and Bob Meet Banach: The Interface of Asymptotic Geometric Analysis and Quantum Information Theory*, volume 223. American Mathematical Soc., 2017.
- [4] Afonso S Bandeira, Yutong Chen, and Amit Singer. Non-unique games over compact groups and orientation estimation in cryo-em. *arXiv preprint arXiv:1505.03840*, 2015.
- [5] Francisco Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):3241, 1982.
- [6] Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 428–437. IEEE, 2016.
- [7] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014.
- [8] Ahron Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. Siam, 2001.
- [9] John J Benedetto and Matthew Fickus. Finite normalized tight frames. *Advances in Computational Mathematics*, 18(2-4):357–385, 2003.
- [10] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge University Press, 2017.

- [11] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012.
- [12] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [13] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.
- [14] Peter J Cameron. 6-transitive graphs. *Journal of Combinatorial Theory, Series B*, 28(2):168–179, 1980.
- [15] Peter G Casazza and Gitta Kutyniok. *Finite frames: Theory and applications*. Springer, 2012.
- [16] Peter G Casazza, Dan Redmond, and Janet C Tremain. Real equiangular frames. In *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*, pages 715–720. IEEE, 2008.
- [17] Lin Chen and Dragomir Ž Djoković. Qubit-qudit states with positive partial transpose. *Physical Review A*, 86(6):062332, 2012.
- [18] Maria Chudnovsky and Paul D Seymour. The structure of claw-free graphs. *Surveys in combinatorics*, 327:153–171, 2005.
- [19] Pete L Clark. Quadratic Reciprocity II: The Proofs (Lecture Notes), 2009. URL: <http://math.uga.edu/~pete/NT2009qrproof.pdf>. Last visited on 2018/05/21. Quotation: “Working through this proof feels a little bit like being an accountant who has been assigned to carefully document a miracle”.
- [20] Charles J Colbourn and Jeffrey H Dinitz. *Handbook of combinatorial designs*. CRC press, 2006.
- [21] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory*, pages 523–562, 2015.
- [22] M.M. Deza and M. Laurent. *Geometry of Cuts and Metrics*. Algorithms and Combinatorics. Springer Berlin Heidelberg, 2009.
- [23] Hamza Fawzi and Pablo A Parrilo. Self-scaled bounds for atomic cone ranks: applications to nonnegative rank and cp-rank. *Mathematical Programming*, 158(1-2):417–465, 2016.
- [24] Hamza Fawzi, James Saunderson, and Pablo A Parrilo. Sparse sums of squares on finite abelian groups and improved semidefinite lifts. *Mathematical Programming*, 160(1-2):149–191, 2016.
- [25] David G Feingold, Richard S Varga, et al. Block diagonally dominant matrices and generalizations of the gerschgorin circle theorem.
- [26] Matthew Fickus, John Jasper, Dustin G Mixon, Jesse D Peterson, and Cody E Watson. Equiangular tight frames with centroidal symmetry. *Applied and Computational Harmonic Analysis*, 2016.

- [27] Matthew Fickus and Dustin G Mixon. Tables of the existence of equiangular tight frames. *arXiv preprint arXiv:1504.00253*, 2015.
- [28] Matthew Fickus and Cody E Watson. Detailing the equivalence between real equiangular tight frames and certain strongly regular graphs. In *Wavelets and Sparsity XVI*, volume 9597, page 959719. International Society for Optics and Photonics, 2015.
- [29] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
- [30] Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *computational complexity*, 10(2):139–154, 2001.
- [31] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613–622, 2001.
- [32] Dima Grigoriev, Edward A Hirsch, and Dmitrii V Pasechnik. Complexity of semi-algebraic proofs. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 419–430. Springer, 2002.
- [33] Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1956.
- [34] Roland Hildebrand. Positive partial transpose from spectra. *Physical Review A*, 76(5):052325, 2007.
- [35] Samuel B Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. *ACM Transactions on Algorithms (TALG)*, 14(3):28, 2018.
- [36] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 720–731. IEEE, 2017.
- [37] M Horedecki, P Horodecki, and R Horodecki. Separability of mixed states: necessary and sufficient conditions phys. *Lett. A*, 223:1–8, 1996.
- [38] Gregg Jaeger. *Quantum information*. Springer, 2007.
- [39] Nathaniel Johnston and Everett Patterson. The inverse eigenvalue problem for entanglement witnesses. *Linear Algebra and its Applications*, 550:1–27, 2018.
- [40] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.
- [41] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan ODonnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007.
- [42] Subhash Khot and Assaf Naor. Grothendieck-type inequalities in combinatorial optimization. *arXiv preprint arXiv:1108.2464*, 2011.

- [43] Subhash Khot and Nisheeth K Vishnoi. On the unique games conjecture. In *FOCS*, volume 5, page 3, 2005.
- [44] Vladimír Kučera. The matrix equation $ax+xb=c$. *SIAM Journal on Applied Mathematics*, 26(1):15–25, 1974.
- [45] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
- [46] Monique Laurent. Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Mathematics of operations research*, 28(4):871–883, 2003.
- [47] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, pages 157–270. Springer, 2009.
- [48] Monique Laurent and Svatopluk Poljak. On a positive semidefinite relaxation of the cut polytope. *Linear Algebra and its Applications*, 223:439–461, 1995.
- [49] Monique Laurent and Svatopluk Poljak. On the facial structure of the set of correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 17(3):530–547, 1996.
- [50] Jon Magne Leinaas, Jan Myrheim, and Per Øyvind Sollid. Numerical studies of entangled positive-partial-transpose states in composite quantum systems. *Physical Review A*, 81(6):062329, 2010.
- [51] Petrus WH Lemmens, Johan J Seidel, and JA Green. Equiangular lines. In *Geometry and Combinatorics*, pages 127–145. Elsevier, 1991.
- [52] Maciej Lewenstein, B Kraus, JI Cirac, and P Horodecki. Optimization of entanglement witnesses. *Physical Review A*, 62(5):052310, 2000.
- [53] Chi-Kwong Li and Bit-Shun Tam. A note on extreme correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 15(3):903–908, 1994.
- [54] Ya-Feng Liu. Set-completely-positive representations and cuts for the max-cut polytope and the unit modulus lifting.
- [55] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 87–96. ACM, 2015.
- [56] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 814–827. ACM, 2016.
- [57] Yurii Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software*, 9(1-3):141–160, 1998.
- [58] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013.
- [59] Gábor Pataki. On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of operations research*, 23(2):339–358, 1998.

- [60] Asher Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8):1413, 1996.
- [61] Amelia Perry, Alexander S Wein, Afonso S Bandeira, Ankur Moitra, et al. Optimality and suboptimality of pca i: Spiked random matrix models. *The Annals of Statistics*, 46(5):2416–2451, 2018.
- [62] Svatopluk Poljak and Zsolt Tuza. Maximum cuts and large bipartite subgraphs. 1995.
- [63] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 245–254. ACM, 2008.
- [64] Prasad Raghavendra, Tselil Schramm, and David Steurer. High-dimensional estimation via sum-of-squares proofs. *arXiv preprint arXiv:1807.11419*, 2018.
- [65] Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 593–602. IEEE, 2008.
- [66] Johan Jacob Seidel. Strongly regular graphs with $(1, 1, 0)$ adjacency matrix having eigenvalue 3. In *Geometry and Combinatorics*, pages 26–43. Elsevier, 1991.
- [67] Mátyás A Sustik, Joel A Tropp, Inderjit S Dhillon, and Robert W Heath Jr. On the existence of equiangular tight frames. *Linear Algebra and its applications*, 426(2-3):619–635, 2007.
- [68] Stanisław J Szarek, Ingemar Bengtsson, and Karol Życzkowski. On the structure of the body of states with positive partial transpose. *Journal of Physics A: Mathematical and General*, 39(5):L119, 2006.
- [69] Luca Trevisan. On Khot’s Unique Games Conjecture. *Bulletin (New Series) of the American Mathematical Society*, 49(1), 2012.
- [70] Lloyd Welch. Lower bounds on the maximum cross correlation of signals (corresp.). *IEEE Transactions on Information theory*, 20(3):397–399, 1974.

A Pseudomoment Reductions for Sum-of-Squares over $\{\pm 1\}^N$

In this appendix, we explain some standard reductions for the degree d SOS relaxation of the problem

$$\mathbf{M}(\mathbf{W}) = \max_{\mathbf{x} \in \{\pm 1\}^N} \mathbf{x}^\top \mathbf{W} \mathbf{x} = \max_{\substack{\mathbf{x} \in \mathbb{R}^N \\ x_i^2 - 1 = 0 \text{ for } i \in [N]}} \sum_{i=1}^N \sum_{j=1}^N W_{ij} x_i x_j. \quad (106)$$

The second expression above writes $\mathbf{M}(\mathbf{W})$ as a polynomial optimization problem, so the standard machinery of SOS optimization (see e.g. [45, 47]) may be applied to formulate the degree d relaxation. We first describe the decision variable of this relaxation.

Definition A.1. *Let d be an even positive integer. Then, $M^{(d)} \subset \mathbb{R}^{N^{\leq d/2} \times N^{\leq d/2}}$ is the set of degree d complete pseudomoment matrices, consisting of \mathbf{Z} whose row and column indices we identify with the set $[N]^{\leq d/2}$ ordered first by ascending length and then lexicographically and satisfying the following properties.*

1. $\mathbf{Z} \succeq \mathbf{0}$.
2. $Z_{\mathbf{st}}$ depends only on $\text{odd}(\mathbf{s} \circ \mathbf{t})$.
3. $Z_{\mathbf{st}} = 1$ whenever $\text{odd}(\mathbf{s} \circ \mathbf{t}) = \emptyset$.

We then define the usual formulation of the degree d SOS relaxation of $\mathbf{M}(\mathbf{W})$ in the following way.

Definition A.2. *Let d be an even positive integer. The degree d SOS relaxation of $\mathbf{M}(\mathbf{W})$ is the optimization problem*

$$\text{SOS}_d(\mathbf{W}) := \max_{\mathbf{Z} \in M^{(d)}} \sum_{i=1}^N \sum_{j=1}^N W_{ij} Z_{(i)(j)}, \quad (107)$$

where $(i) \in [N]$ and $(j) \in [N]$ are interpreted as strings of length 1.

The result we will prove in this appendix is that the pseudomoment matrices of Definition A.1 can be truncated to just the minor indexed by $[N]^{d/2} \times [N]^{d/2}$ without affecting the optimization problem (107). First, we make the simple observation that, by Condition 2 of Definition A.1, the objective function of $\text{SOS}_d(\mathbf{W})$ may be rewritten in terms of this minor.

Definition A.3. *For any $N \geq 1$, let $\mathbf{e}_k \in [N]^k$ be the string of length k with all entries equal to the symbol $1 \in [N]$.*

Proposition A.4. *For each d an even positive integer,*

$$\text{SOS}_d(\mathbf{W}) = \max_{\mathbf{Z} \in M^{(d)}} \sum_{i=1}^N \sum_{j=1}^N W_{ij} Z_{(\mathbf{e}_{d/2-1} \circ (i))(\mathbf{e}_{d/2-1} \circ (j))}. \quad (108)$$

We next show that it does not matter whether we define the set of minors of $\mathbf{Z} \in M^{(d)}$ indexed by $[N]^{d/2} \times [N]^{d/2}$ by truncating \mathbf{Z} or by applying the constraints of Definition A.1 to only a subset of strings $[N]^{d/2} \subset [N]^{\leq d/2}$, as we did in the main text in Definition 2.3.

Definition A.5. *Let d be an even positive integer. Then, $\tilde{M}^{(d)} \subset \mathbb{R}^{[N]^{d/2} \times [N]^{d/2}}$ is the set of degree d truncated pseudomoment matrices, consisting of $\tilde{\mathbf{Z}}$ satisfying the properties of Definition A.1 but only for strings of length exactly $d/2$, that is, for $\mathbf{s}, \mathbf{t} \in [N]^{d/2}$.*

Proposition A.6. $\tilde{M}^{(d)}$ is equal to the set of $\tilde{\mathbf{Z}}$ occurring as the minor indexed by $[N]^{d/2} \times [N]^{d/2}$ of $\mathbf{Z} \in M^{(d)}$.

Proof. Clearly if $\mathbf{Z} \in M^{(d)}$ then the $[N]^{d/2} \times [N]^{d/2}$ minor of \mathbf{Z} will belong to $\tilde{M}^{(d)}$. Therefore, it suffices to prove that for any $\tilde{\mathbf{Z}} \in \tilde{M}^{(d)}$, there exists $\mathbf{Z} \in M^{(d)}$ such that $\tilde{\mathbf{Z}}$ is the $[N]^{d/2} \times [N]^{d/2}$ minor of \mathbf{Z} .

We define the entries of \mathbf{Z} to be

$$Z_{\mathbf{s}\mathbf{t}} := \tilde{Z}_{(\mathbf{s} \circ \mathbf{e}_{d/2-|\mathbf{s}|})(\mathbf{t} \circ \mathbf{e}_{d/2-|\mathbf{t}|})}. \quad (109)$$

(Intuitively, this construction is in analogy to the possibility of assuming without loss of generality that $x_1 = 1$ in the optimization defining $\mathbf{M}(\mathbf{W})$.) By construction, $\tilde{\mathbf{Z}}$ is the necessary minor of \mathbf{Z} . We have

$$\text{odd}((\mathbf{s} \circ \mathbf{e}_{d/2-|\mathbf{s}|}) \circ (\mathbf{t} \circ \mathbf{e}_{d/2-|\mathbf{t}|})) = \text{odd}(\mathbf{s} \circ \mathbf{t} \circ \mathbf{e}_{|\mathbf{s}+\mathbf{t}|}), \quad (110)$$

and $|\mathbf{s}| + |\mathbf{t}| \equiv |\text{odd}(\mathbf{s} \circ \mathbf{t})| \pmod{2}$, so $Z_{\mathbf{s}\mathbf{t}}$ is a function of only $\text{odd}(\mathbf{s} \circ \mathbf{t})$. Also, if $\text{odd}(\mathbf{s} \circ \mathbf{t}) = \emptyset$ then $|\mathbf{s}| + |\mathbf{t}|$ must be even, so in this case the expression in (110) also equals \emptyset , thus in this case $Z_{\mathbf{s}\mathbf{t}} = 1$ since $\tilde{\mathbf{Z}} \in \tilde{M}^{(d)}$.

It then only remains to show that $\mathbf{Z} \succeq \mathbf{0}$ to show that $\mathbf{Z} \in M^{(d)}$. For two strings $\mathbf{s}, \mathbf{s}' \in [N]^{<\infty}$, let us write $\mathbf{s} \leq_1 \mathbf{s}'$ if $|\mathbf{s}| \leq |\mathbf{s}'|$ and $\mathbf{s}' = \mathbf{s} \circ \mathbf{e}_{|\mathbf{s}'|-|\mathbf{s}|}$. Suppose that $\mathbf{v} \in \mathbb{R}^{[N]^{\leq d/2}}$, then using the above definition we may write its quadratic form with \mathbf{Z} as

$$\begin{aligned} \mathbf{v}^\top \mathbf{Z} \mathbf{v} &= \sum_{\mathbf{s} \in [N]^{\leq d/2}} \sum_{\mathbf{t} \in [N]^{\leq d/2}} Z_{\mathbf{s}\mathbf{t}} v_{\mathbf{s}} v_{\mathbf{t}} \\ &= \sum_{\mathbf{s}' \in [N]^{d/2}} \sum_{\mathbf{t}' \in [N]^{d/2}} \tilde{Z}_{\mathbf{s}'\mathbf{t}'} \left(\sum_{\substack{\mathbf{s} \in [N]^{\leq d/2} \\ \mathbf{s} \leq_1 \mathbf{s}'}} \sum_{\substack{\mathbf{t} \in [N]^{\leq d/2} \\ \mathbf{t} \leq_1 \mathbf{t}'}} v_{\mathbf{s}} v_{\mathbf{t}} \right) \\ &= \sum_{\mathbf{s}' \in [N]^{d/2}} \sum_{\mathbf{t}' \in [N]^{d/2}} \tilde{Z}_{\mathbf{s}'\mathbf{t}'} \left(\sum_{\substack{\mathbf{s} \in [N]^{\leq d/2} \\ \mathbf{s} \leq_1 \mathbf{s}'}} v_{\mathbf{s}} \right) \left(\sum_{\substack{\mathbf{t} \in [N]^{\leq d/2} \\ \mathbf{t} \leq_1 \mathbf{t}'}} v_{\mathbf{t}} \right) \\ &\geq 0, \end{aligned} \quad (111)$$

where the last inequality follows because $\tilde{\mathbf{Z}} \succeq \mathbf{0}$. Thus, $\mathbf{Z} \in M^{(d)}$, completing the proof. \square

The result we were interested in then follows, that $\text{SOS}_d(\mathbf{W})$ may equivalently be defined in terms of optimization over the truncated pseudomoment matrices $\tilde{M}^{(d)}$.

Corollary A.7. For each d an even positive integer,

$$\max_{\tilde{\mathbf{Z}} \in \tilde{M}^{(d)}} \sum_{i=1}^N \sum_{j=1}^N W_{ij} \tilde{Z}_{(\mathbf{e}_{d/2-1} \circ (i))(\mathbf{e}_{d/2-1} \circ (j))}. \quad (112)$$

B Proofs of Structural Results on $\mathcal{B}(N, r)$

B.1 Proof of Proposition 4.1

Let $\mathbf{M} \in \mathcal{B}(N, r)$. To obtain the spectral bound on the blocks $\|\mathbf{M}_{[ij]}\| \leq 1$, note that the claim is trivial for $i = j$, so let us fix $i, j \in [N]$ with $i \neq j$ and denote $\mathbf{S} := \mathbf{M}_{[ij]} \in \mathbb{R}_{\text{sym}}^{r \times r}$. Taking a suitable

minor of \mathbf{M} , we find

$$\begin{bmatrix} \mathbf{I}_r & \mathbf{S} \\ \mathbf{S} & \mathbf{I}_r \end{bmatrix} \succeq \mathbf{0}. \quad (113)$$

Taking a quadratic form with this matrix, we find that for any $\mathbf{v} \in \mathbb{R}^r$ with $\|\mathbf{v}\|_2 = 1$,

$$0 \leq \begin{bmatrix} \pm \mathbf{v} \\ \mathbf{v} \end{bmatrix}^\top \begin{bmatrix} \mathbf{I}_r & \mathbf{S} \\ \mathbf{S} & \mathbf{I}_r \end{bmatrix} \begin{bmatrix} \pm \mathbf{v} \\ \mathbf{v} \end{bmatrix} = 2 \pm 2\mathbf{v}^\top \mathbf{S} \mathbf{v}, \quad (114)$$

thus $|\mathbf{v}^\top \mathbf{S} \mathbf{v}| \leq 1$, and the result follows.

From this, the bound $\|\mathbf{M}\| \leq N$ follows from a simple case of the ‘‘block Gershgorin circle theorem’’ [25], which may be deduced directly as follows: suppose $\mathbf{v} \in \mathbb{R}^{rN}$ is the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$, then

$$\mathbf{v}^\top \mathbf{M} \mathbf{v} \leq \sum_{i=1}^N \sum_{j=1}^N |\mathbf{v}_i^\top \mathbf{M}_{[ij]} \mathbf{v}_j| \leq \sum_{i=1}^N \sum_{j=1}^N \|\mathbf{v}_i\|_2 \|\mathbf{v}_j\|_2 = \left(\sum_{i=1}^N \|\mathbf{v}_i\|_2 \right)^2 \leq N \sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N \|\mathbf{v}\|_2^2, \quad (115)$$

giving the result.

For the final statement of the Proposition, if $\mathbf{M} \mathbf{v} = N \mathbf{v}$, then all of the inequalities in (115) must be equalities. For the third inequality to be an equality requires all of the $\|\mathbf{v}_i\|_2$ to be equal for $i \in [N]$. For the first inequality to be an equality requires $\mathbf{v}_i^\top \mathbf{M}_{[ij]} \mathbf{v}_j \geq 0$ for all $i, j \in [N]$. For the second inequality to be an equality requires $\mathbf{M}_{[ij]} \mathbf{v}_j = \mathbf{v}_i$ for all $i, j \in [N]$, completing the proof.

B.2 Proof of Proposition 4.2

Let $\mathbf{M} \in \mathcal{B}(N, r)$ and let $r' := \text{rank}(\mathbf{M})$. Since \mathbf{M} contains \mathbf{I}_r as a minor, $r' \geq r$, and since rN is the dimension of \mathbf{M} , $r' \leq rN$. Then, there exists $\mathbf{U} \in \mathbb{R}^{r' \times rN}$ such that $\mathbf{M} = \mathbf{U}^\top \mathbf{U}$. Let us expand in blocks

$$\mathbf{U} = [\mathbf{U}_1 \quad \mathbf{U}_2 \quad \cdots \quad \mathbf{U}_N], \quad (116)$$

for $\mathbf{U}_i \in \mathbb{R}^{r' \times r}$. Then, $\mathbf{U}_i^\top \mathbf{U}_i = \mathbf{M}_{[ii]} = \mathbf{I}_r$.

This factorization is unchanged by multiplying \mathbf{U} on the left by any matrix of $\mathcal{O}(r')$. Since \mathbf{U}_1 has orthogonal columns, by choosing a suitable such multiplication we may assume without loss of generality that the columns of \mathbf{U}_1 are the first r standard basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_r \in \mathbb{R}^{r'}$. Equivalently,

$$\mathbf{U}_1 = \left[\begin{array}{c} \mathbf{I}_r \\ \mathbf{0} \end{array} \right] \begin{array}{l} \} r \\ \} r' - r \end{array}. \quad (117)$$

Let us expand each \mathbf{U}_i in blocks of the same dimensions,

$$\mathbf{U}_i =: \left[\begin{array}{c} \mathbf{S}_i \\ \mathbf{R}_i \end{array} \right] \begin{array}{l} \} r \\ \} r' - r \end{array}, \quad (118)$$

then $\mathbf{S}_1 = \mathbf{I}_r$ and $\mathbf{R}_1 = \mathbf{0}$. We first show that the \mathbf{S}_i are all symmetric. Expanding the block $\mathbf{M}_{[1i]}$, we have

$$\mathbf{M}_{[1i]} = \mathbf{U}_1^\top \mathbf{U}_i = \mathbf{S}_1^\top \mathbf{S}_i + \mathbf{R}_1^\top \mathbf{R}_i = \mathbf{S}_i, \quad (119)$$

and since $\mathbf{M}_{[1i]}$ is symmetric, \mathbf{S}_i is symmetric as well.

It remains to show the relations (51) and (52). For the former, we expand $\mathbf{M}_{[ii]}$:

$$\mathbf{I}_r = \mathbf{M}_{[ii]} = \mathbf{U}_i^\top \mathbf{U}_i = \mathbf{S}_i^2 + \mathbf{R}_i^\top \mathbf{R}_i. \quad (120)$$

For the latter, we expand $M_{[ij]}$ and $M_{[ji]}$:

$$\mathbf{0} = M_{[ij]} - M_{[ji]} = \mathbf{U}_i^\top \mathbf{U}_j - \mathbf{U}_j^\top \mathbf{U}_i = \mathbf{S}_i \mathbf{S}_j - \mathbf{S}_j \mathbf{S}_i + \mathbf{R}_i^\top \mathbf{R}_j - \mathbf{R}_j^\top \mathbf{R}_i. \quad (121)$$

C Proofs of Relaxation Descriptions of Theorem 2.15

C.1 Proof of Proposition 4.3

Positive direction. Suppose $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$, $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$, $\mathbf{v} \in \mathbb{R}^{rN}$ is the concatenation of $\mathbf{v}_1, \dots, \mathbf{v}_N$, $\sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N$, and $\mathbf{M} \in \mathcal{B}(N, r)$ with $\text{rank}(\mathbf{M}) = r$ and $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$. By Proposition 4.1, $\|\mathbf{v}_i\|_2 = 1$ for each $i \in [N]$ and $M_{[ij]} \mathbf{v}_j = \mathbf{v}_i$ for each $i, j \in [N]$.

Since $\mathbf{M} \succeq 0$ and $\text{rank}(\mathbf{M}) = r$, there exist $\mathbf{Q}_i \in \mathbb{R}^{r \times r}$ such that $M_{[ij]} = \mathbf{Q}_i^\top \mathbf{Q}_j$. Moreover, since $\mathbf{Q}_i^\top \mathbf{Q}_i = M_{[ii]} = \mathbf{I}_r$, $\mathbf{Q}_i \in \mathcal{O}(r)$ for each $i \in [N]$. The above factorization is unchanged by multiplying each \mathbf{Q}_i on the left by an orthogonal matrix, so we may assume without loss of generality that $\mathbf{Q}_1 = \mathbf{I}_r$.

Thus, $M_{[1i]} = \mathbf{Q}_1^\top \mathbf{Q}_i = \mathbf{Q}_i$, which must be symmetric, so \mathbf{Q}_i is symmetric for each $i \in [N]$. And, $M_{[ij]} = \mathbf{Q}_i \mathbf{Q}_j$ is also symmetric, so $\mathbf{Q}_1, \dots, \mathbf{Q}_N$ are a commuting family of symmetric orthogonal matrices. Therefore, there exists some $\mathbf{Q} \in \mathcal{O}(r)$ and $\mathbf{1} = \mathbf{d}_1, \dots, \mathbf{d}_N \in \{\pm 1\}^r$ such that $\mathbf{Q}_i = \mathbf{Q} \mathbf{D}_i \mathbf{Q}^\top$ where $\mathbf{D}_i = \text{diag}(\mathbf{d}_i)$.

We have $\mathbf{v}_i = M_{[i1]} \mathbf{v}_1 = \mathbf{Q}_i \mathbf{v}_1 = \mathbf{Q} \mathbf{D}_i \mathbf{Q}^\top \mathbf{v}_1$ for each $i \in [N]$. Thus,

$$X_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \langle \mathbf{D}_i \mathbf{Q}^\top \mathbf{v}_1, \mathbf{D}_j \mathbf{Q}^\top \mathbf{v}_1 \rangle = \langle \mathbf{D}_i \mathbf{D}_j, \mathbf{Q}^\top \mathbf{v}_1 \mathbf{v}_1^\top \mathbf{Q} \rangle. \quad (122)$$

Let $\boldsymbol{\rho} = \text{diag}(\mathbf{Q}^\top \mathbf{v}_1 \mathbf{v}_1^\top \mathbf{Q})$, then since $\mathbf{Q}^\top \mathbf{v}_1 \mathbf{v}_1^\top \mathbf{Q} \succeq \mathbf{0}$, $\boldsymbol{\rho} \geq 0$, and $\sum_{i=1}^r \rho_i = \text{Tr}(\mathbf{Q}^\top \mathbf{v}_1 \mathbf{v}_1^\top \mathbf{Q}) = 1$. Therefore, letting $\tilde{\mathbf{d}}_k := ((\mathbf{d}_i)_k)_{i=1}^N \in \{\pm 1\}^N$, (122) is

$$X_{ij} = \sum_{k=1}^r \rho_k (\mathbf{d}_i)_k (\mathbf{d}_j)_k, \quad (123)$$

$$\mathbf{X} = \sum_{k=1}^r \rho_k \tilde{\mathbf{d}}_k \tilde{\mathbf{d}}_k^\top \in \mathcal{C}^N, \quad (124)$$

completing the proof.

Negative direction. We have $\mathbf{I}_N \in \mathcal{C}^N$ since $\mathbf{I}_N = \frac{1}{2^N} \sum_{\mathbf{x} \in \{\pm 1\}^N} \mathbf{x} \mathbf{x}^\top$, as each off-diagonal entry occurs an equal number of times with a positive sign as with a negative sign in the summation. We will view $\mathbf{I}_N = \text{Gram}(\mathbf{e}_1, \dots, \mathbf{e}_N)$, let $\mathbf{v} = \sum_{i=1}^N \mathbf{e}_i \otimes \mathbf{e}_i$ be the concatenation of the \mathbf{e}_i , and will show that if $\mathbf{M} \in \mathcal{B}(N, N)$ with $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$, then $\text{rank}(\mathbf{M}) > N$ when $N \notin \{1, 2\} \cup 4\mathbb{N}$.

Suppose otherwise, then, as in the argument above, $\mathbf{M} \in \mathcal{B}(N, N)$ has $M_{[ij]} = \mathbf{Q}_i \mathbf{Q}_j$ for some $\mathbf{Q}_i \in \mathcal{O}(N) \cap \mathbb{R}_{\text{sym}}^{N \times N}$, with $\mathbf{Q}_1 = \mathbf{I}_N$, and where $\mathbf{Q}_1, \dots, \mathbf{Q}_N$ commute. We may then write $\mathbf{Q}_i = \mathbf{Q} \mathbf{D}_i \mathbf{Q}^\top$ for $\mathbf{Q} \in \mathcal{O}(N)$ and $\mathbf{D}_i = \text{diag}(\mathbf{d}_i)$ for $\mathbf{d}_i \in \{\pm 1\}^N$. Let us also write $\mathbf{q}_1, \dots, \mathbf{q}_N$ for the rows of \mathbf{Q} , which form an orthonormal basis of \mathbb{R}^N .

We have

$$N^2 = \mathbf{v}^\top \mathbf{M} \mathbf{v} = \sum_{i=1}^N \sum_{j=1}^N (\mathbf{e}_i \otimes \mathbf{e}_i)^\top \mathbf{M} (\mathbf{e}_j \otimes \mathbf{e}_j) = \sum_{i=1}^N \sum_{j=1}^N (M_{[ij]})_{ij}. \quad (125)$$

Since $\mathbf{M} \succeq 0$ and $\text{diag}(\mathbf{M}) = \mathbf{1}$, all entries of \mathbf{M} are at most 1, so each term in this sum must equal 1, i.e. $(M_{[ij]})_{ij} = 1$ for all $i, j \in [N]$. We then have, for any i, j ,

$$1 = (M_{[ij]})_{ij} = \mathbf{e}_i^\top \mathbf{Q} \mathbf{D}_i \mathbf{D}_j \mathbf{Q}^\top \mathbf{e}_j = \langle \mathbf{D}_i \mathbf{q}_i, \mathbf{D}_j \mathbf{q}_j \rangle, \quad (126)$$

whereby $\mathbf{D}_i \mathbf{q}_i = \mathbf{D}_j \mathbf{q}_j$ for all i, j . In other words, there exists some $\mathbf{q} \in \mathbb{R}^N$ with $\|\mathbf{q}\|_2 = 1$ such that $\mathbf{D}_i \mathbf{q}_i = \mathbf{q}$, or $\mathbf{q}_i = \mathbf{D}_i \mathbf{q}$. Thus, the \mathbf{q}_i are sign flips of a fixed vector.

On the other hand, the \mathbf{q}_i are the rows of $\mathbf{Q} \in \mathcal{O}(N)$, whose columns must also form an orthonormal basis. Therefore, every entry of \mathbf{q} must have the same norm, so each entry of \mathbf{Q} also has equal norm; in other words, \mathbf{Q} is, up to a scaling depending on definitions, a Hadamard matrix with real entries [20]. A real-valued Hadamard matrix of order N can only exist when $N \in \{1, 2\} \cup 4\mathbb{N}$, so under the assumptions of the Proposition this is a contradiction.

This example is simple to analyze but probably suboptimal. In general, a suitable example for this result is a matrix $\mathbf{X} \in \mathcal{C}^N$ where $\text{rank}(\mathbf{X})$ is strictly smaller than the smallest number of cut matrices $\mathbf{x}_1 \mathbf{x}_1^\top, \dots, \mathbf{x}_m \mathbf{x}_m^\top$ for $\mathbf{x}_i \in \{\pm 1\}^N$ such that $\mathbf{X} \in \text{conv}(\{\mathbf{x}_i \mathbf{x}_i^\top\}_{i=1}^m)$. The latter quantity is similar to the notions of *completely-positive rank* and *non-negative rank*, and appears to behave counterintuitively sometimes; see [23, 54] for some discussion.

C.2 Proof of Proposition 4.5

Suppose first that $\mathbf{X} = \text{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N)$ for $\mathbf{v}_i \in \mathbb{R}^r$ with $\sum_{i=1}^N \|\mathbf{v}_i\|_2^2 = N$, and $\mathbf{M} \in \mathcal{B}_{\text{sep}}(N, r)$ such that $\mathbf{v}^\top \mathbf{M} \mathbf{v} = N^2$. By Proposition 4.1, $\|\mathbf{v}_i\|_2 = 1$ for each $i \in [N]$. By absorbing constants and rearranging tensor products, the condition $\mathbf{M} \in \mathcal{B}_{\text{sep}}(N, r)$ may be rewritten as

$$\mathbf{M} = \sum_{i=1}^m \mathbf{A}_i \otimes (\mathbf{b}_i \mathbf{b}_i^\top) \quad (127)$$

for some $\mathbf{A}_i \in \mathbb{R}_{\text{sym}}^{N \times N}$ with $\mathbf{A}_i \succeq \mathbf{0}$ and such that, letting $\mathbf{a}_i = \text{diag}(\mathbf{A}_i)$,

$$\sum_{i=1}^m (\mathbf{a}_i)_j \mathbf{b}_i \mathbf{b}_i^\top = \mathbf{I}_r \quad (128)$$

for each $j \in [N]$.

Let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns. Then,

$$\mathbf{v}^\top \mathbf{M} \mathbf{v} = \sum_{i=1}^m \sum_{j=1}^N \sum_{k=1}^N (\mathbf{A}_i)_{jk} \langle \mathbf{b}_i, \mathbf{v}_j \rangle \langle \mathbf{b}_i, \mathbf{v}_k \rangle = \sum_{i=1}^m \mathbf{b}_i^\top \mathbf{V} \mathbf{A}_i \mathbf{V}^\top \mathbf{b}_i. \quad (129)$$

We now bound $\mathbf{b}_i^\top \mathbf{V} \mathbf{A}_i \mathbf{V}^\top \mathbf{b}_i$ by applying a simple matrix inequality; the rather complicated formulation below is only to handle carefully the possibility of certain diagonal entries of \mathbf{A}_i equaling zero. Let $\tilde{\mathbf{A}}_i$ be the maximal strictly positive definite minor of \mathbf{A}_i , of dimension N_i , and let \mathbf{w}_i be the restriction of $\mathbf{V}^\top \mathbf{b}_i$ to the same indices. Then, $\text{diag}(\tilde{\mathbf{A}}_i) > 0$. Let $\pi_i : [N_i] \rightarrow [N]$ map the indices of this minor to the original indices, and let us define a diagonal matrix $\mathbf{D}_i \in \mathbb{R}^{N_i \times N_i}$ by

$$(\mathbf{D}_i)_{jj} := \left(\sum_{j'=1}^{N_i} \sqrt{(\tilde{\mathbf{A}}_i)_{j'j'}} \cdot |\langle \mathbf{b}_i, \mathbf{v}_{\pi_i(j')} \rangle| \right) \frac{|\langle \mathbf{b}_i, \mathbf{v}_{\pi_i(j)} \rangle|}{\sqrt{(\tilde{\mathbf{A}}_i)_{jj}}}. \quad (130)$$

Then, we claim $\mathbf{D}_i \succeq \mathbf{w}_i \mathbf{w}_i^\top$. This is a matter of applying a weighted Cauchy-Schwarz inequality:

for $\mathbf{x} \in \mathbb{R}^N$, we have

$$\begin{aligned}
\mathbf{x}^\top \mathbf{w}_i \mathbf{w}_i^\top \mathbf{x} &= \left(\sum_{j=1}^{N_i} x_j \langle \mathbf{b}_i, \mathbf{v}_{\pi_i(j)} \rangle \right)^2 \\
&\leq \left(\sum_{j'=1}^{N_i} \sqrt{(\tilde{\mathbf{A}}_i)_{j'j'}} \cdot |\langle \mathbf{b}_i, \mathbf{v}_{\pi_i(j')} \rangle| \right) \left(\sum_{j=1}^N \frac{|\langle \mathbf{b}_i, \mathbf{v}_{\pi_i(j)} \rangle|}{\sqrt{(\tilde{\mathbf{A}}_i)_{jj}}} x_j^2 \right) \\
&= \sum_{j=1}^N (\mathbf{D}_i)_{jj} x_j^2.
\end{aligned} \tag{131}$$

Therefore,

$$\begin{aligned}
\mathbf{b}_i^\top \mathbf{V} \mathbf{A}_i \mathbf{V}^\top \mathbf{b}_i &= \mathbf{w}_i^\top \tilde{\mathbf{A}}_i \mathbf{w}_i \\
&\leq \langle \mathbf{D}_i, \tilde{\mathbf{A}}_i \rangle \\
&= \left(\sum_{j=1}^{N_i} \sqrt{(\tilde{\mathbf{A}}_i)_{jj}} \cdot |\langle \mathbf{b}_i, \mathbf{v}_{\pi_i(j)} \rangle| \right)^2 \\
&= \left(\sum_{j=1}^N \sqrt{(\mathbf{a}_i)_j} \cdot |\langle \mathbf{b}_i, \mathbf{v}_j \rangle| \right)^2.
\end{aligned} \tag{132}$$

Now, combining (132) with (128) and (129) and using the Cauchy-Schwarz inequality, we find

$$\mathbf{v}^\top \mathbf{M} \mathbf{v} \leq N \sum_{i=1}^m \sum_{j=1}^N (\mathbf{a}_i)_j \langle \mathbf{b}_i, \mathbf{v}_j \rangle^2 = N \sum_{j=1}^N \|\mathbf{v}_j\|_2^2 = N^2. \tag{133}$$

Thus, the Cauchy-Schwarz inequality in (133) must be tight, whereby there exist $\kappa_i \geq 0$ with $\sum_{i=1}^m \kappa_i = 1$ such that

$$(\mathbf{a}_i)_j \langle \mathbf{b}_i, \mathbf{v}_j \rangle^2 = \kappa_i \tag{134}$$

for every $i \in [m]$ and $j \in [N]$. Note in particular that if $\kappa_i > 0$ for some $i \in [m]$, then $\langle \mathbf{b}_i, \mathbf{v}_j \rangle \neq 0$ for all $j \in [N]$. We may then define vectors $\boldsymbol{\beta}_{jk} \in \mathbb{R}^m$ by

$$(\boldsymbol{\beta}_{jk})_i := \begin{cases} \sqrt{\kappa_i} \frac{\langle \mathbf{b}_i, \mathbf{v}_j \rangle}{\langle \mathbf{b}_i, \mathbf{v}_k \rangle} & : \kappa_i > 0, \\ 0 & : \kappa_i = 0. \end{cases} \tag{135}$$

Then,

$$\begin{aligned}
\|\boldsymbol{\beta}_{jk}\|_2^2 &= \sum_{i:\kappa_i>0} \kappa_i \frac{\langle \mathbf{b}_i, \mathbf{v}_j \rangle^2}{\langle \mathbf{b}_i, \mathbf{v}_k \rangle^2} \\
&= \sum_{i:\kappa_i>0} (\mathbf{a}_i)_k \langle \mathbf{b}_i, \mathbf{v}_j \rangle^2 \\
&\leq \sum_{i=1}^m (\mathbf{a}_i)_k \langle \mathbf{b}_i, \mathbf{v}_j \rangle^2 \\
&\stackrel{(128)}{=} 1,
\end{aligned} \tag{136}$$

$$\langle \boldsymbol{\beta}_{jk}, \boldsymbol{\beta}_{kj} \rangle = \sum_{i:\kappa_i>0} \kappa_i = 1. \tag{137}$$

Thus, in fact $\|\beta_{jk}\|_2 = 1$ and $\beta_{jk} = \beta_{kj}$ for all $j, k \in [N]$. This implies first that whenever $\kappa_i > 0$ then $\langle \mathbf{b}_i, \mathbf{v}_j \rangle^2$ does not depend on j , and second that whenever $\kappa_i = 0$ then $(\mathbf{a}_i)_k \langle \mathbf{b}_i, \mathbf{v}_j \rangle^2 = 0$ for all $j, k \in [N]$. We may assume without loss of generality that $\mathbf{A}_i \neq \mathbf{0}$, so $\mathbf{a}_i \neq \mathbf{0}$, and thus the latter implies that whenever $\kappa_i = 0$, then $\langle \mathbf{b}_i, \mathbf{v}_j \rangle^2 = 0$ for all $j \in [N]$. Therefore, in all cases, $\langle \mathbf{b}_i, \mathbf{v}_j \rangle^2$ does not depend on j .

Let us write $\eta_i := \langle \mathbf{b}_i, \mathbf{v}_j \rangle^2$. For i where $\eta_i \neq 0$, by (134) $(\mathbf{a}_i)_j$ does not depend on j either. For these i , let us write $\phi_i := (\mathbf{a}_i)_j$. Evaluating (128) as a bilinear form on \mathbf{v}_j and \mathbf{v}_k , we then find

$$X_{jk} = \langle \mathbf{v}_j, \mathbf{v}_k \rangle = \sum_{i:\eta_i \neq 0} \phi_i \langle \mathbf{b}_i, \mathbf{v}_j \rangle \langle \mathbf{b}_i, \mathbf{v}_k \rangle = \sum_{i:\eta_i \neq 0} \phi_i \eta_i \operatorname{sgn}(\langle \mathbf{b}_i, \mathbf{v}_j \rangle) \operatorname{sgn}(\langle \mathbf{b}_i, \mathbf{v}_k \rangle). \quad (138)$$

When $\eta_i \neq 0$, then $\phi_i \eta_i = \kappa_i$, and when $\eta_i = 0$ then $\kappa_i = 0$. Therefore, we have in fact

$$X_{jk} = \sum_{i=1}^m \kappa_i \operatorname{sgn}(\langle \mathbf{b}_i, \mathbf{v}_j \rangle) \operatorname{sgn}(\langle \mathbf{b}_i, \mathbf{v}_k \rangle), \quad (139)$$

showing $\mathbf{X} \in \mathcal{C}^N$.

The converse is simpler: suppose that $\mathbf{X} \in \mathcal{C}^N$ and $\mathbf{X} = \operatorname{Gram}(\mathbf{v}_1, \dots, \mathbf{v}_N) \in \mathcal{C}^N$ for $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{R}^r$. Let $\mathbf{v} \in \mathbb{R}^{rN}$ be the concatenation of the $\mathbf{v}_1, \dots, \mathbf{v}_N$. We will build $\mathbf{M} \in \mathcal{B}_{\text{sep}}(N, r)$ by essentially reversing the process described in the proof of Proposition 4.3. Let $\rho_1, \dots, \rho_m \geq 0$ with $\sum_{i=1}^m \rho_i = 1$ and $\tilde{\mathbf{d}}_1, \dots, \tilde{\mathbf{d}}_m \in \{\pm 1\}^N$ be such that

$$\mathbf{X} = \sum_{k=1}^m \rho_k \tilde{\mathbf{d}}_k \tilde{\mathbf{d}}_k^\top. \quad (140)$$

We may assume without loss of generality that $m \geq r$, by adding extra terms with zero coefficient to this expression. Then, writing $\mathbf{d}_i := ((\tilde{\mathbf{d}}_k)_i)_{k=1}^m \in \mathbb{R}^m$, $\mathbf{R} = \operatorname{diag}(\boldsymbol{\rho})$, and $\mathbf{v}'_i = \mathbf{R}^{1/2} \mathbf{d}_i$, (140) implies that $\mathbf{X} = \operatorname{Gram}(\mathbf{v}'_1, \dots, \mathbf{v}'_m)$. There then exists $\mathbf{Z} \in \mathbb{R}^{m \times r}$ such that $\mathbf{Z} \mathbf{v}_i = \mathbf{v}'_i$ and $\mathbf{Z}^\top \mathbf{Z} = \mathbf{I}_r$.

We let $\mathbf{D}_i := \operatorname{diag}(\mathbf{d}_i)$, and define $\mathbf{M} \in \mathbb{R}^{rN \times rN}$ to have blocks

$$\mathbf{M}_{[ij]} := \mathbf{Z}^\top \mathbf{D}_i \mathbf{D}_j \mathbf{Z} = (\mathbf{D}_i \mathbf{Z})^\top (\mathbf{D}_j \mathbf{Z}). \quad (141)$$

The last expression gives \mathbf{M} as a Gram matrix, so $\mathbf{M} \succeq \mathbf{0}$. Since $\mathbf{D}_i^2 = \mathbf{I}_r$ for each $i \in [N]$, $\mathbf{M}_{[ii]} = \mathbf{I}_r$, and since $\mathbf{D}_1, \dots, \mathbf{D}_N$ commute, $\mathbf{M}_{[ij]}$ is symmetric. Thus, $\mathbf{M} \in \mathcal{B}(N, r)$. We also have

$$\mathbf{v}^\top \mathbf{M} \mathbf{v} = \sum_{i=1}^N \sum_{j=1}^N \mathbf{v}'_i{}^\top \mathbf{D}_i \mathbf{D}_j \mathbf{v}'_j{}^\top = \sum_{i=1}^N \sum_{j=1}^N \mathbf{d}_i{}^\top \mathbf{R}^{1/2} \mathbf{D}_i \mathbf{D}_j \mathbf{R}^{1/2} \mathbf{d}_j = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^m \rho_k = N^2. \quad (142)$$

It only remains to check that \mathbf{M} is separable. To do this, let $\mathbf{z}_1, \dots, \mathbf{z}_m \in \mathbb{R}^r$ be the rows of \mathbf{Z} , then it is straightforward to check against (141) that we can write $\mathbf{M} = \sum_{i=1}^m (\tilde{\mathbf{d}}_i \otimes \mathbf{z}_i) (\tilde{\mathbf{d}}_i \otimes \mathbf{z}_i)^\top$.

D Proofs of Results on Partial Transposition

D.1 Proof of Proposition 5.1

This result is simply a matter of applying the vectorization operation vec to the singular value decomposition: if $\mathbf{V} = \sum_{i=1}^r \sigma_i \mathbf{y}_i \mathbf{z}_i^\top$ for $\mathbf{y}_i \in \mathbb{R}^r$ and $\mathbf{z}_i \in \mathbb{R}^N$, then, noting that $\operatorname{vec}(\mathbf{y}_i \mathbf{z}_i^\top) = \mathbf{z}_i \otimes \mathbf{y}_i$ and $\operatorname{vec} : \mathbb{R}^{r \times N} \rightarrow \mathbb{R}^{rN}$ is linear, the result follows.

D.2 Proof of Proposition 5.2

Suppose $\mathbf{V} \in \mathbb{R}^{r \times N}$ with $r \leq N$ has singular value decomposition $\mathbf{V} = \sum_{i=1}^r \sigma_i \mathbf{y}_i \mathbf{z}_i^\top$ for orthonormal sets of $\mathbf{y}_i \in \mathbb{R}^r$ and $\mathbf{z}_i \in \mathbb{R}^N$ and with $\sigma_i \geq 0$. Let $\mathbf{v} = \text{vec}(\mathbf{V})$. Applying Proposition 5.1, we may write

$$\begin{aligned} \mathbf{v} \mathbf{v}^\top &= \left(\sum_{i=1}^r \sigma_i \mathbf{z}_i \otimes \mathbf{y}_i \right) \left(\sum_{i=1}^r \sigma_i \mathbf{z}_i \otimes \mathbf{y}_i \right)^\top \\ &= \sum_{i=1}^r \sum_{j=1}^r \sigma_i \sigma_j (\mathbf{z}_i \mathbf{z}_j^\top) \otimes (\mathbf{y}_i \mathbf{y}_j^\top) \\ &= \sum_{i=1}^r \sum_{j=1}^r \sigma_i \sigma_j (\mathbf{z}_i \otimes \mathbf{y}_i) (\mathbf{z}_j \otimes \mathbf{y}_j)^\top. \end{aligned} \quad (143)$$

Therefore, the partial transpose is

$$\begin{aligned} (\mathbf{v} \mathbf{v}^\top)^\Gamma &= \sum_{i=1}^r \sum_{j=1}^r \sigma_i \sigma_j (\mathbf{z}_i \mathbf{z}_j^\top) \otimes (\mathbf{y}_j \mathbf{y}_i^\top) \\ &= \sum_{i=1}^r \sum_{j=1}^r \sigma_i \sigma_j (\mathbf{z}_i \otimes \mathbf{y}_j) \otimes (\mathbf{z}_j \otimes \mathbf{y}_i)^\top \\ &= \sum_{i=1}^r \sigma_i^2 (\mathbf{z}_i \otimes \mathbf{y}_i) (\mathbf{z}_i \otimes \mathbf{y}_i)^\top \\ &\quad + \sum_{1 \leq i < j \leq r} \sigma_i \sigma_j \left((\mathbf{z}_i \otimes \mathbf{y}_j) (\mathbf{z}_j \otimes \mathbf{y}_i)^\top + (\mathbf{z}_j \otimes \mathbf{y}_i) (\mathbf{z}_i \otimes \mathbf{y}_j)^\top \right), \end{aligned} \quad (144)$$

and the result follows by diagonalizing the rank-two matrices in the second sum.

D.3 Proof of Proposition 5.3

Suppose $\mathbf{V} \in \mathbb{R}^{r \times N}$ with $r \leq N$ has full rank and singular value decomposition $\mathbf{V} = \sum_{i=1}^r \sigma_i \mathbf{y}_i \mathbf{z}_i^\top$ for an orthonormal basis of $\mathbf{y}_i \in \mathbb{R}^r$ and an orthonormal set $\mathbf{z}_i \in \mathbb{R}^N$, with $\sigma_i > 0$ by the full-rank condition. Let $\mathbf{v} = \text{vec}(\mathbf{V})$. Let us also extend $\mathbf{z}_1, \dots, \mathbf{z}_r$ with $\mathbf{z}_{r+1}, \dots, \mathbf{z}_N$ to a full orthonormal basis.

Since $\mathbf{V} \mathbf{V}^\top = \sum_{i=1}^r \sigma_i^2 \mathbf{y}_i \mathbf{y}_i^\top$, we may expand

$$\mathbf{I}_N \otimes (\mathbf{V} \mathbf{V}^\top) = \left(\sum_{i=1}^N \mathbf{z}_i \mathbf{z}_i^\top \right) \otimes \left(\sum_{j=1}^r \sigma_j^2 \mathbf{y}_j \mathbf{y}_j^\top \right) = \sum_{i=1}^N \sum_{j=1}^r \sigma_j^2 (\mathbf{z}_i \otimes \mathbf{y}_j) (\mathbf{z}_i \otimes \mathbf{y}_j)^\top. \quad (145)$$

Dividing this sum into those summands with $i \leq r$ and those with $i > r$ and subtracting (144), we

may write

$$\begin{aligned}
& \mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma \\
&= \sum_{1 \leq i < j \leq r} \left(\frac{1}{2} \sigma_i^2 (\mathbf{z}_j \otimes \mathbf{y}_i) (\mathbf{z}_j \otimes \mathbf{y}_i)^\top + \frac{1}{2} \sigma_j^2 (\mathbf{z}_i \otimes \mathbf{y}_j) (\mathbf{z}_i \otimes \mathbf{y}_j)^\top \right. \\
&\quad \left. - \sigma_i \sigma_j (\mathbf{z}_i \otimes \mathbf{y}_j) (\mathbf{z}_j \otimes \mathbf{y}_i)^\top - \sigma_i \sigma_j (\mathbf{z}_j \otimes \mathbf{y}_i) (\mathbf{z}_i \otimes \mathbf{y}_j)^\top \right) \\
&\quad + \sum_{i=r+1}^N \sum_{j=1}^r \sigma_j^2 (\mathbf{z}_i \otimes \mathbf{y}_j) (\mathbf{z}_i \otimes \mathbf{y}_j)^\top \\
&= \frac{1}{2} \sum_{1 \leq i < j \leq r} (\sigma_i \mathbf{z}_j \otimes \mathbf{y}_i - \sigma_j \mathbf{z}_i \otimes \mathbf{y}_j) (\sigma_i \mathbf{z}_j \otimes \mathbf{y}_i - \sigma_j \mathbf{z}_i \otimes \mathbf{y}_j)^\top \\
&\quad + \sum_{i=r+1}^N \sum_{j=1}^r \sigma_j^2 (\mathbf{z}_i \otimes \mathbf{y}_j) (\mathbf{z}_i \otimes \mathbf{y}_j)^\top. \tag{146}
\end{aligned}$$

We thus find an alternate proof that $\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma \succeq \mathbf{0}$. The benefit of this approach is that it allows us to read off the subspace we are interested in directly: note that up to rescaling the expression (146) is a spectral decomposition, and thus

$$\begin{aligned}
& \ker \left(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma \right)^\perp \\
&= \text{span} \left(\left\{ \frac{1}{\sqrt{\sigma_i^2 + \sigma_j^2}} (\sigma_i \mathbf{z}_j \otimes \mathbf{y}_i - \sigma_j \mathbf{z}_i \otimes \mathbf{y}_j) \right\}_{1 \leq i < j \leq r} \cup \{ \mathbf{z}_i \otimes \mathbf{y}_j \}_{i \in [N] \setminus [r], j \in [r]} \right), \tag{147}
\end{aligned}$$

$$\begin{aligned}
& \ker \left(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma \right) \\
&= \text{span} \left(\left\{ \frac{1}{\sqrt{\sigma_i^2 + \sigma_j^2}} (\sigma_i \mathbf{z}_j \otimes \mathbf{y}_i + \sigma_j \mathbf{z}_i \otimes \mathbf{y}_j) \right\}_{1 \leq i < j \leq r} \cup \{ \mathbf{z}_i \otimes \mathbf{y}_i \}_{i \in [r]} \right), \tag{148}
\end{aligned}$$

where the first equality follows from (146) and the second may be checked by counting dimensions and verifying mutual orthogonalities. It is also straightforward to verify that the vectors enumerated in (148) are orthonormal, and thus give an orthonormal basis for $\ker(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma)$.

The only remaining task is to check the alternate description

$$\ker \left(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma \right) \stackrel{?}{=} \{ \text{vec}(\mathbf{S}\mathbf{V}) : \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r} \} =: V_{\text{sym}}. \tag{149}$$

We have $\dim(\ker(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma)) = \frac{r(r+1)}{2}$ by (148). Since \mathbf{v}_i are a spanning set, if $\text{vec}(\mathbf{S}\mathbf{V}) = \mathbf{0}$ then $\mathbf{S} = \mathbf{0}$, so the map $\mathbf{S} \mapsto \text{vec}(\mathbf{S}\mathbf{V})$ is injective and thus $\dim(V_{\text{sym}}) = \dim(\mathbb{R}_{\text{sym}}^{r \times r}) = \frac{r(r+1)}{2}$ as well. Therefore, to show (149) it suffices to show one inclusion.

Suppose that $\mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r}$, then

$$\begin{aligned}
((\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma) \text{vec}(\mathbf{S}\mathbf{V}))_{[i]} &= (\mathbf{V}\mathbf{V}^\top) \mathbf{S}\mathbf{v}_i - \sum_{j=1}^N \mathbf{v}_j \mathbf{v}_i^\top \mathbf{S}\mathbf{v}_j \\
&= \sum_{j=1}^N \mathbf{v}_j \mathbf{v}_j^\top \mathbf{S}\mathbf{v}_i - \sum_{j=1}^N \mathbf{v}_j \mathbf{v}_i^\top \mathbf{S}\mathbf{v}_j \\
&= \mathbf{0},
\end{aligned} \tag{150}$$

where in the last step we use that \mathbf{S} is symmetric. Thus, $\text{vec}(\mathbf{S}\mathbf{V}) \in \ker(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma)$, so $V_{\text{sym}} \subseteq \ker(\mathbf{I}_N \otimes (\mathbf{V}\mathbf{V}^\top) - (\mathbf{v}\mathbf{v}^\top)^\Gamma)$, which completes the proof by the previous dimension counting argument.

E Tight Frame Projector Calculations

In this appendix we will derive formulae for the orthogonal projectors to various subspaces associated with a UNTF $\mathbf{v}_1, \dots, \mathbf{v}_N \in \mathbb{S}^{r-1}$, and the specializations to the case of ETFs. Let $\mathbf{V} \in \mathbb{R}^{r \times N}$ have the \mathbf{v}_i as its columns, then recall that we define a map $\mathcal{V} : \mathbb{R}_{\text{sym}}^{r \times r} \rightarrow \mathbb{R}^{rN}$ by $\mathcal{V}(\mathbf{S}) = \sqrt{\frac{r}{N}} \text{vec}(\mathbf{S}\mathbf{V})$, which by Proposition 6.2 is a linear isometric embedding. Let us also write $\mathbf{X} = \mathbf{V}^\top \mathbf{V} \in \mathbb{R}^{N \times N}$ for the Gram matrix.

We then are interested in the projector to the following subspace:

$$V'_{\text{sym}} := \mathcal{V} \left(\left\{ \mathbf{S} \in \mathbb{R}_{\text{sym}}^{r \times r} : \langle \mathbf{S}, \mathbf{v}_i \mathbf{v}_i^\top \rangle = 0 \text{ for } i \in [N] \right\} \right). \tag{151}$$

As a warmup, we will also consider the following simpler subspace:

$$V_{\text{sym}} := \mathcal{V}(\mathbb{R}_{\text{sym}}^{r \times r}). \tag{152}$$

The idea of the calculation in both cases will be as follows: suppose $V \subset \mathbb{R}_{\text{sym}}^{r \times r}$ is some subspace and $\mathbf{y} \in \mathbb{R}^{rN}$ is the concatenation of $\mathbf{y}_1, \dots, \mathbf{y}_N \in \mathbb{R}^r$. Then by the variational characterization of the orthogonal projector, $\mathbf{P}_{\mathcal{V}(V)} \mathbf{y} = \mathcal{V}(\mathbf{S}^*(\mathbf{y}))$, where

$$\begin{aligned}
\text{obj}(\mathbf{S}; \mathbf{y}) &:= \frac{1}{2} \sum_{i=1}^N \left\| \sqrt{\frac{r}{N}} \mathbf{S}\mathbf{v}_i - \mathbf{y}_i \right\|_2^2 \\
&= \frac{1}{2} \|\mathbf{y}\|_2^2 + \frac{1}{2} \text{Tr}(\mathbf{S}^2) - \left\langle \mathbf{S}, \sqrt{\frac{r}{N}} \sum_{i=1}^N \frac{\mathbf{v}_i \mathbf{y}_i^\top + \mathbf{y}_i \mathbf{v}_i^\top}{2} \right\rangle,
\end{aligned} \tag{153}$$

$$\mathbf{S}^*(\mathbf{y}) = \text{argmin}_{\mathbf{S} \in V} \text{obj}(\mathbf{S}; \mathbf{y}). \tag{154}$$

a minimization which we will solve by introducing Lagrange multipliers for the constraint $\mathbf{S} \in V$, which will reduce the task to solving a linear system in the Lagrange multiplier variables.⁴

⁴Note that the simple form of the quadratic term in \mathbf{S} is a consequence of the \mathbf{v}_i forming a UNTF, whereby $\sum_{i=1}^N \mathbf{v}_i \mathbf{v}_i^\top = \mathbf{V}\mathbf{V}^\top = \frac{N}{r} \mathbf{I}_r$. In a more general setting, the matrix $\mathbf{V}\mathbf{V}^\top$ would appear and, upon differentiating with respect to \mathbf{S} , we would not get a formula for the optimizer \mathbf{S}^* but rather a so-called *continuous matrix Lyapunov equation* $(\mathbf{V}\mathbf{V}^\top) \mathbf{S}^* + \mathbf{S}^* (\mathbf{V}\mathbf{V}^\top) = \mathbf{Q}$. Such an equation in principle admits an analytic solution by reducing to a linear equation in $\text{vec}(\mathbf{S}^*)$ (see e.g. [44]), but this would further complicate the calculations.

E.1 Projector to V_{sym}

We illustrate the first part of this idea below for the simplest case of $V = V_{\text{sym}}$, where no Lagrange multipliers are required.

Proposition E.1. $P_{V_{\text{sym}}} = \frac{r}{2N}(\mathbf{X} \otimes \mathbf{I}_r + (\mathbf{v}\mathbf{v}^\top)^\Gamma)$, where $\mathbf{X} = \mathbf{V}^\top \mathbf{V}$ is the Gram matrix of the \mathbf{v}_i .

Proof. We will solve the variational description (154) with $V = V_{\text{sym}}$. Since the optimization is unconstrained, we may compute directly the first-order condition for the optimizer

$$0 = \frac{\partial \text{obj}}{\partial \mathbf{S}}(\mathbf{S}^*; \mathbf{y}) = \mathbf{S}^* - \sqrt{\frac{r}{N}} \sum_{j=1}^N \frac{\mathbf{v}_j \mathbf{y}_j^\top + \mathbf{y}_j \mathbf{v}_j^\top}{2}, \quad (155)$$

thus the optimizer is

$$\mathbf{S}^* = \mathbf{S}^*(\mathbf{y}) = \sqrt{\frac{r}{N}} \sum_{j=1}^N \frac{\mathbf{v}_j \mathbf{y}_j^\top + \mathbf{y}_j \mathbf{v}_j^\top}{2}. \quad (156)$$

Then, the blocks of the projection of \mathbf{y} may be recovered as

$$\begin{aligned} (\mathbf{P}_{V_{\text{sym}}} \mathbf{y})_{[i]} &= (\mathcal{V}(\mathbf{S}^*))_{[i]} \\ &= \sqrt{\frac{r}{N}} \mathbf{S}^* \mathbf{v}_i \\ &= \frac{r}{2N} \sum_{j=1}^N (\langle \mathbf{y}_j, \mathbf{v}_i \rangle \mathbf{v}_j + \langle \mathbf{v}_i, \mathbf{v}_j \rangle \mathbf{y}_j) \\ &= \sum_{j=1}^N \frac{r}{2N} (\langle \mathbf{v}_i, \mathbf{v}_j \rangle \mathbf{I}_r + \mathbf{v}_j \mathbf{v}_i^\top) \mathbf{y}_j. \end{aligned} \quad (157)$$

In particular, the matrices in each term of the sum give the blocks $(\mathbf{P}_{V_{\text{sym}}})_{[ij]}$, whereby the formula in the statement is clearly correct for each block. \square

E.2 Projector to V'_{sym}

The case of $V = V'_{\text{sym}}$, being a constrained optimization, requires more intermediate calculations in order to determine the Lagrange multipliers. An important role will be played by the Hadamard square of the Gram matrix, $\mathbf{X}^{\odot 2}$, which is equivalently the Gram matrix of the matrices $\mathbf{v}_i \mathbf{v}_i^\top$ (and which figured in the proof of the Gerzon bound, Proposition 6.3, and the description of perturbations of matrices in the elliptope, Proposition 5.4). To perform our calculations in closed form, we will need to compute the inverse of this matrix explicitly. We will first give a general result in terms of this inverse and then show how the inverse may be computed for the case of ETFs needed in the main text.

Proposition E.2. *Suppose that the matrices $\mathbf{v}_i \mathbf{v}_i^\top$ are linearly independent, or equivalently that the matrix $\mathbf{X}^{\odot 2}$ is non-singular. Then, the blocks of $\mathbf{P}_{V'_{\text{sym}}}$ are given by*

$$(\mathbf{P}_{V'_{\text{sym}}})_{[ij]} = \frac{r}{N} \left(\frac{1}{2} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \mathbf{I}_r + \frac{1}{2} \mathbf{v}_j \mathbf{v}_i^\top - \sum_{k=1}^N \sum_{\ell=1}^N ((\mathbf{X}^{\odot 2})^{-1})_{k\ell} \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_\ell \rangle \mathbf{v}_k \mathbf{v}_\ell^\top \right). \quad (158)$$

Proof. We must now solve (154) with $T = V'_{\text{sym}} = \text{span}(\{\mathbf{v}_i \mathbf{v}_i^\top : i \in [N]\})^\perp$. We introduce the Lagrangian

$$L(\mathbf{S}, \boldsymbol{\gamma}; \mathbf{y}) := \text{obj}(\mathbf{S}; \mathbf{y}) - \left\langle \mathbf{S}, \sum_{i=1}^N \gamma_i \mathbf{v}_i \mathbf{v}_i^\top \right\rangle \quad (159)$$

and write the first-order condition $\frac{\partial L}{\partial \mathbf{S}}(\mathbf{S}^*, \boldsymbol{\gamma}; \mathbf{y}) = 0$, which gives

$$\mathbf{S}^* = \mathbf{S}^*(\mathbf{y}) = \sqrt{\frac{r}{N}} \sum_{j=1}^N \frac{\mathbf{v}_j \mathbf{y}_j^\top + \mathbf{y}_j \mathbf{v}_j^\top}{2} + \sum_{j=1}^N \gamma_j \mathbf{v}_j \mathbf{v}_j^\top. \quad (160)$$

The other first-order condition $\frac{\partial L}{\partial \boldsymbol{\gamma}}(\mathbf{S}^*, \boldsymbol{\gamma}; \mathbf{y}) = 0$ is equivalent to the constraints, $\langle \mathbf{S}^*, \mathbf{v}_i \mathbf{v}_i^\top \rangle = 0$ for all $i \in [N]$, which yields the system of linear equations for $\boldsymbol{\gamma}$,

$$\sum_{j=1}^N (\mathbf{X}^{\odot 2})_{ij} \gamma_j = -\sqrt{\frac{r}{N}} \sum_{j=1}^N \langle \mathbf{v}_i, \mathbf{v}_j \rangle \langle \mathbf{v}_i, \mathbf{y}_j \rangle \text{ for } i \in [N]. \quad (161)$$

Since $\mathbf{X}^{\odot 2}$ is invertible by assumption, this admits a unique solution which is given by

$$\gamma_j = -\sqrt{\frac{r}{N}} \sum_{k=1}^N \sum_{\ell=1}^N ((\mathbf{X}^{\odot 2})^{-1})_{jk} \langle \mathbf{v}_k, \mathbf{v}_\ell \rangle \langle \mathbf{v}_k, \mathbf{y}_\ell \rangle. \quad (162)$$

Substituting into (160), we find

$$\mathbf{S}^* = \sqrt{\frac{r}{N}} \left(\sum_{j=1}^N \frac{\mathbf{v}_j \mathbf{y}_j^\top + \mathbf{y}_j \mathbf{v}_j^\top}{2} - \sum_{j=1}^N \sum_{k=1}^N \sum_{\ell=1}^N ((\mathbf{X}^{\odot 2})^{-1})_{jk} \langle \mathbf{v}_k, \mathbf{v}_\ell \rangle \langle \mathbf{v}_k, \mathbf{y}_\ell \rangle \mathbf{v}_j \mathbf{v}_j^\top \right). \quad (163)$$

As before, we recover the blocks of the projection of \mathbf{y} ,

$$\begin{aligned} (\mathbf{P}_{V'_{\text{sym}}} \mathbf{y})_{[i]} &= (\mathcal{V}(\mathbf{S}^*))_{[i]} \\ &= \sqrt{\frac{r}{N}} \mathbf{S}^* \mathbf{v}_i \\ &= \frac{r}{N} \left(\sum_{j=1}^N \frac{\langle \mathbf{v}_i, \mathbf{y}_j \rangle \mathbf{v}_j + \langle \mathbf{v}_i, \mathbf{v}_j \rangle \mathbf{y}_j}{2} - \sum_{j=1}^N \sum_{k=1}^N \sum_{\ell=1}^N ((\mathbf{X}^{\odot 2})^{-1})_{jk} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \langle \mathbf{v}_k, \mathbf{v}_\ell \rangle \langle \mathbf{v}_k, \mathbf{y}_\ell \rangle \mathbf{v}_j \right) \\ &= \sum_{j=1}^N \frac{r}{N} \left(\frac{1}{2} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \mathbf{I}_r + \frac{1}{2} \mathbf{v}_j \mathbf{v}_i^\top - \sum_{k=1}^N \sum_{\ell=1}^N ((\mathbf{X}^{\odot 2})^{-1})_{k\ell} \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_\ell \rangle \mathbf{v}_k \mathbf{v}_\ell^\top \right) \mathbf{y}_j, \end{aligned} \quad (164)$$

and the result follows. \square

Corollary E.3. *Suppose that $\mathbf{v}_1, \dots, \mathbf{v}_N$ form an ETF with $r > 1$. Then, the blocks of $\mathbf{P}_{V'_{\text{sym}}}$ are given by*

$$(\mathbf{P}_{V'_{\text{sym}}})_{[ij]} = \frac{N-r}{N(r-1)} \mathbf{v}_i \mathbf{v}_j^\top + \frac{r}{2N} \mathbf{v}_j \mathbf{v}_i^\top + \frac{r}{2N} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \mathbf{I}_r - \frac{r^2(N-1)}{N^2(r-1)} \sum_{k=1}^N \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_k \rangle \mathbf{v}_k \mathbf{v}_k^\top. \quad (165)$$

Proof. By Proposition 6.3, the conditions of Proposition E.2 are satisfied, so it suffices to compute $(\mathbf{X}^{\odot 2})^{-1}$. The off-diagonal entries of \mathbf{X} all equal the coherence α , which by Proposition 2.13 is given by

$$\alpha = \sqrt{\frac{N-r}{r(N-1)}}. \quad (166)$$

Thus, we have

$$\mathbf{X}^{\odot 2} = (1 - \alpha^2)\mathbf{I}_N + \alpha^2\mathbf{1}\mathbf{1}^\top = \frac{N(r-1)}{r(N-1)}\mathbf{I}_N + \frac{N-r}{r(N-1)}\mathbf{1}\mathbf{1}^\top. \quad (167)$$

This matrix may be inverted by the Sherman-Morrison formula, giving

$$(\mathbf{X}^{\odot 2})^{-1} = \frac{r(N-1)}{N(r-1)}\mathbf{I}_N - \frac{r(N-r)}{N^2(r-1)}\mathbf{1}\mathbf{1}^\top. \quad (168)$$

Thus, the entries are

$$(\mathbf{X}^{\odot 2})_{ij}^{-1} = \begin{cases} a := \frac{r((N-1)^2+r-1)}{N^2(r-1)} & : i = j, \\ b := -\frac{r(N-r)}{N^2(r-1)} & : i \neq j. \end{cases} \quad (169)$$

Substituting into the expression from Proposition E.2, we find

$$\begin{aligned} & \sum_{k=1}^N \sum_{\ell=1}^N ((\mathbf{X}^{\odot 2})^{-1})_{k\ell} \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_\ell \rangle \mathbf{v}_k \mathbf{v}_\ell^\top \\ &= (a-b) \sum_{k=1}^N \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_k \rangle \mathbf{v}_k \mathbf{v}_k^\top + b \sum_{k=1}^N \sum_{\ell=1}^N \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_\ell \rangle \mathbf{v}_k \mathbf{v}_\ell^\top \\ &= (a-b) \sum_{k=1}^N \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_k \rangle \mathbf{v}_k \mathbf{v}_k^\top + b(\mathbf{V}\mathbf{V}^\top \mathbf{v}_i)(\mathbf{V}\mathbf{V}^\top \mathbf{v}_j)^\top \\ &= \frac{r(N-1)}{N(r-1)} \sum_{k=1}^N \langle \mathbf{v}_i, \mathbf{v}_k \rangle \langle \mathbf{v}_j, \mathbf{v}_k \rangle \mathbf{v}_k \mathbf{v}_k^\top - \frac{N-r}{r(r-1)} \mathbf{v}_i \mathbf{v}_j^\top. \end{aligned} \quad (170)$$

Combining with the full result of Proposition E.2 then gives the claim. \square