GRASSMANNIAN OPTIMIZATION IS NP-HARD

ZEHUA LAI, LEK-HENG LIM, AND KE YE

Abstract. We show that unconstrained quadratic optimization over a Grassmannian $\mathrm{Gr}(k, n)$ is NP-hard. Our results cover all scenarios: (i) when k and n are both allowed to grow; (ii) when k is arbitrary but fixed; (iii) when k is fixed at its lowest possible value 1. We then deduce the NP-hardness of unconstrained cubic optimization over the Stiefel manifold $\mathrm{V}(k, n)$ and the orthogonal group $\mathrm{O}(n)$. As an addendum we demonstrate the NP-hardness of unconstrained quadratic optimization over the Cartan manifold, i.e., the positive definite cone $\mathbb{S}^n_{++}$ regarded as a Riemannian manifold, another popular example in manifold optimization. We will also establish the nonexistence of FPTAS in all cases.

1. Introduction

It is well-known that unconstrained quadratic programming is easy over the n-space $\mathbb{R}^n$ or n-sphere S^n. The former is a simple exercise[1] whereas the latter is polynomial-time solvable [42, Section 4.3]. It is therefore somewhat surprising that optimizing a quadratic function over the Grassmannian $\mathrm{Gr}(k, n)$ is an NP-hard problem, not least because the $k = 1$ case is intimately related to $\mathbb{R}^n$ and S^n: $\mathrm{Gr}(1, n + 1) = \mathbb{RP}^n$ is a compactification of $\mathbb{R}^n$ and S^n is a double-cover of $\mathrm{Gr}(1, n + 1)$. We will prove three variations of this result:

ASSUMPTION	REDUCTION	SECTION
k and n can both grow	clique decision	Section 4
k arbitrary but fixed, only n can grow	clique number	Section 5
$k = 1$, only n can grow	stability number	Section 7

We first establish these using the representation $\mathrm{Gr}(k, n) = \{P \in \mathbb{R}^{n\times n} : P^2 = P = P^{\mathsf{T}},\ \mathrm{tr}(P) = k\}$, and then extend them to all known models of the Grassmannian.

These hardness results for the Grassmannian form the mainstay of our article, accounting for the choice of title. But by leveraging on them, we will also deduce the NP-hardness of optimization problems over some of the most common manifolds in applied mathematics:

MANIFOLD	HARDNESS RESULT	REDUCTION
Stiefel manifold $\mathrm{V}(k, n)$	unconstrained cubic programming is NP-hard	QP over $\mathrm{Gr}(k, n)$
Orthogonal group $\mathrm{O}(n)$	unconstrained cubic programming is NP-hard	CP over $\mathrm{V}(k, n)$
Cartan manifold $\mathbb{S}^n_{++}$	unconstrained quadratic programming is NP-hard	matrix copositivity

Here the Stiefel manifold is modeled as the set of $n \times k$ matrices with orthonormal columns [40], and the Cartan manifold is modeled as the positive definite cone $\mathbb{S}^n_{++}$ but equipped with the Riemannian metric $\mathsf{g}_A(X, Y) = \mathrm{tr}(A^{-1}XA^{-1}Y)$ for $A \in \mathbb{S}^n_{++}$ [3]. Note that as $\mathrm{V}(1, n+1) = \mathrm{S}^n$ and unconstrained quadratic programming on the n-sphere is polynomial-time, we need at least cubic programming (CP) to demonstrate NP-hardnesss over $\mathrm{V}(k, n)$ for all possible k, n.

Notations. We write S^n for the n-sphere and $\mathbb{S}^n$ for the set of $n \times n$ symmetric matrices. We write $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ for the positive semidefinite and positive definite cones respectively. The operation $\mathrm{diag} : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ takes a matrix X to its vector of diagonal entries $\mathrm{diag}(X)$. We write $\mathrm{O}(n) :=$

[1] Minimum of $\frac{1}{2}x^{\mathsf{T}}Ax - b^{\mathsf{T}}x$ is either attained at $x_* = A^{-1}b$ if A is positive definite or else it is $-\infty$. The common refrain "QP is NP-hard" [41] refers to constrained QP over $\mathbb{R}^n$.

$\{Q \in \mathbb{R}^{n\times n} : Q^{\mathsf{T}}Q = I\}$ for the orthogonal group and $\mathrm{GL}(n) := \{X \in \mathbb{R}^{n\times n} : \operatorname{rank}(X) = n\}$ for the general linear group. A matrix group is a subgroup of $\mathrm{GL}(m)$ for some $m \in \mathbb{N}$. A matrix submanifold is a submanifold of $\mathbb{R}^{m\times n}$ for some $m, n \in \mathbb{N}$ and a matrix quotient manifold is one that is given by the quotient of a matrix submanifold by a matrix group; collectively we refer to both as matrix manifolds. We identify $\mathbb{R}^n = \mathbb{R}^{n\times 1}$, i.e., an $x \in \mathbb{R}^n$ will always be a column vector; row vectors will be written x^{T}. We delimit equivalence classes with double brackets $[\![\,\cdot\,]\!]$.

2. Models of the Grassmann, Stiefel, and Cartan manifolds

As *abstract manifolds*, the Grassmannian is the set of k-dimensional subspaces in $\mathbb{R}^n$, the noncompact Stiefel manifold is the set of k-frames in $\mathbb{R}^n$, the compact Stiefel manifold is the set of orthonormal k-frames in $\mathbb{R}^n$, and the Cartan manifold is the set of ellipsoids in $\mathbb{R}^n$ centered at origin. We denote them by $\mathrm{Gr}_k(\mathbb{R}^n)$, $\mathrm{St}_k(\mathbb{R}^n)$, $\mathrm{V}_k(\mathbb{R}^n)$, $\mathcal{E}(\mathbb{R}^n)$ respectively, following standard notations in geometry and topology, whenever we want to refer to them as abstract manifolds.

To do anything with these manifolds, one needs a concrete *model* that parameterizes points on the manifold, possibly also vectors on tangent spaces, that is convenient to work with. Put in another way, a model endows the manifold with a system of extrinsic coordinates. In the context of this article, without a model it would not even make sense to speak of quadratic or cubic programming, as the degree of a polynomial function invariably depends on the coordinates used to parameterize points on the manifold.

The models favored by practitioners, particularly those in manifold optimization, are typically chosen with a view towards computational feasibility, and basically fall into two distinct families:

(i) *quotient models*: represent points and tangent vectors as equivalence classes of matrices;
(ii) *submanifold models*: represent points and tangent vectors as actual matrices.

As we will see in Section 6, the NP-hardness claimed in the previous section will extend to every model.

The quotient models are matrix quotient manifolds $\mathcal{M}/G$ obtained from a matrix group G acting on a matrix submanifold $\mathcal{M}$ through matrix multiplication. Table 1 lists all quotient models for the Stiefel, Grassmann, and Cartan manifolds that we are aware of:

Table 1. Quotient models

$\mathcal{M}$	G	$\mathcal{M}/G$
$\mathrm{O}(n)$	$\mathrm{O}(k)\times\mathrm{O}(n-k)$	$\mathrm{Gr}_k(\mathbb{R}^n)$
$\mathrm{V}(k,n)$	$\mathrm{O}(k)$	$\mathrm{Gr}_k(\mathbb{R}^n)$
$\mathrm{GL}(n)$	$\mathrm{P}(k,n)$	$\mathrm{Gr}_k(\mathbb{R}^n)$
$\mathrm{St}(k,n)$	$\mathrm{GL}(k)$	$\mathrm{Gr}_k(\mathbb{R}^n)$
$\mathrm{O}(n)$	$\mathrm{O}(n-k)$	$\mathrm{V}_k(\mathbb{R}^n)$
$\mathrm{GL}(n)$	$\mathrm{P}_1(k,n)$	$\mathrm{St}_k(\mathbb{R}^n)$
$\mathrm{GL}(n)$	$\mathrm{O}(n)$	$\mathcal{E}(\mathbb{R}^n)$

The submanifolds $\mathrm{V}(k,n)$ and $\mathrm{St}(k,n)$ are respectively models for the compact and noncompact Stiefel manifolds given in Table 2. The group of 2×2 block upper triangular matrices with invertible diagonal blocks,

$$\mathrm{P}(k,n) := \left\{\begin{bmatrix} X_1 & X_2 \\ 0 & X_3 \end{bmatrix} \in \mathrm{GL}(n) : X_1 \in \mathrm{GL}(k),\ X_2 \in \mathbb{R}^{k\times(n-k)},\ X_3 \in \mathrm{GL}(n-k)\right\}$$

is called a *parabolic subgroup* of $\mathrm{GL}(n)$; and $\mathrm{P}_1(k,n)$ is the subgroup of $\mathrm{P}(k,n)$ with $X_1 = I_k$.

The last column in Table 1 gives the abstract manifold that $\mathcal{M}/G$ is diffeomorphic to but note that $\mathcal{M}/G$ inherits an extrinsic coordinate system given by $\mathcal{M}$ and the G-action that is absent in its abstract counterpart. Some of the manifold in Table 1 are well-known homogeneous spaces and some even symmetric spaces [17] but we will not use these properties. The first two quotient models in Table 1 have formed the mainstay of Grassmannian manifold optimization with tens of thousands of works, too numerous for us to identify any handful of representative articles, and we will just mention the pioneering feat of [11].

The submanifold models are simply matrix submanifolds $\mathcal{M}$ diffeomorphic to one of the abstract manifold in the last column in Table 2, which lists all such models for the Stiefel, Grassmann, and Cartan manifolds known to us:

Table 2. Submanifold models

model	concrete	abstract
positive definite model	$\mathbb{S}^n_{++} = \{A \in \mathbb{R}^{n\times n} : A^{\mathsf{T}} = A,\ x^{\mathsf{T}}Ax > 0 \text{ for all } x \neq 0\}$	$\mathcal{E}(\mathbb{R}^n)$
orthonormal model	$\mathrm{V}(k,n) = \{Y \in \mathbb{R}^{n\times k} : Y^{\mathsf{T}}Y = I\}$	$\mathrm{V}_k(\mathbb{R}^n)$
full-rank model	$\mathrm{St}(k,n) = \{S \in \mathbb{R}^{n\times k} : \operatorname{rank}(S) = k\}$	$\mathrm{St}_k(\mathbb{R}^n)$
projection model	$\mathrm{Gr}_\pi(k,n) = \{P \in \mathbb{R}^{n\times n} : P^2 = P = P^{\mathsf{T}},\ \operatorname{tr}(P) = k\}$	$\mathrm{Gr}_k(\mathbb{R}^n)$
involution model	$\mathrm{Gr}_\iota(k,n) = \{Q \in \mathbb{R}^{n\times n} : Q^{\mathsf{T}}Q = I,\ Q^{\mathsf{T}} = Q,\ \operatorname{tr}(Q) = 2k-n\}$	$\mathrm{Gr}_k(\mathbb{R}^n)$
(a,b)-model	$\mathrm{Gr}_{a,b}(k,n) = \{W \in \mathbb{R}^{n\times n} : W^{\mathsf{T}} = W,\ (W-aI)(W-bI) = 0,$ $\operatorname{tr}(W) = ka + (n-k)b\}$	$\mathrm{Gr}_k(\mathbb{R}^n)$

As in the case of the quotient models, these submanifold models endow the abstract manifold with an extrinsic coordinate system. The (a,b)-model for the $\mathrm{Gr}_k(\mathbb{R}^n)$ is an infinite family of submanifold models, one for each distinct pair of $a,b \in \mathbb{R}$, that has only recently been discovered [26, Theorem 6.1]. It contains both the projection and involution models as special cases:

$$\mathrm{Gr}_{1,0}(k,n) = \mathrm{Gr}_\pi(k,n), \qquad \mathrm{Gr}_{1,-1}(k,n) = \mathrm{Gr}_\iota(k,n).$$

The projection model has a long history in applications including coding theory [2, 6], machine learning [12], optimization [39, 44], statistics [5]; and is also common in geometric measure theory [27] and differential geometry [35]. The involution model is more recent [21, 22, 26].

The submanifold models enjoys an advantage for our goal: It is clear what one means by a polynomial function of degree d, namely, it is one that has degree d in the matrix variables. The same task is considerably more involved for the quotient models and we will not attempt it in this article. Nevertheless we will show in Proposition 6.1 that any two different models in Tables 1 and 2 for the same abstract manifold may be transformed to each other in polynomial-time.

By far the two most common models in pure mathematics are $\mathrm{GL}(n)/\mathrm{P}(k,n)$ in Table 1 and the *Plücker model*. For the sake of completeness the latter is the image of the Plücker embedding $\mathrm{Gr}(k,n) \to \mathbb{RP}^{\binom{n}{k}}$ [16, Lecture 6]. It is hard to imagine it ever finding practical use due to the exponential dimension of its ambient space. We will disregard it in the rest of this article.

3. Complexity theoretic assumptions

For our purpose in this article, it suffices to take an informal approach towards the optimization theoretic terms used. "Unconstrained" simply means that we do not impose any constraints on the optimization variables other than that the points must lie on the manifold $\mathcal{M}$ in question. For submanifold models, we only consider real-valued objective functions $f : \mathcal{M} \to \mathbb{R}$ that are polynomial in the entries of the matrix variable $X \in \mathcal{M} \subseteq \mathbb{R}^{m\times n}$. In fact, every objective function on submanifold models in this article will be a polynomial function $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ that is everywhere defined on $\mathbb{R}^{m\times n}$. By the degree of f, we simply means its degree as a polynomial in the variables x_{ij}'s.

For quotient models we will not use the notion of degree as a rigorous definition is beyond the scope of this article. To see why it can be tricky to define "degree" over a quotient model, take $\mathrm{O}(3)/(\mathrm{O}(2)\times \mathrm{O}(1))$ for illustration. We have

$$\left[\!\!\left[\begin{bmatrix} c & 0 & -s \\ 0 & 1 & 0 \\ s & 0 & c \end{bmatrix}\right]\!\!\right] = \left[\!\!\left[\begin{bmatrix} c^3 - cs^2 & -2c^2 s & -s \\ 2cs & c^2 - s^2 & 0 \\ c^2 s - s^3 & -2cs^2 & c \end{bmatrix}\right]\!\!\right] \tag{1}$$

where $[\![\,\cdot\,]\!]$ denotes $\mathrm{O}(2)\times\mathrm{O}(1)$-cosets in $\mathrm{O}(3)$. In fact the matrix representative on the right can be chosen to have arbitrary high powers in c and s, noting that it is really just the $d = 2$ case of

$$\begin{bmatrix} c & 0 & -s \\ 0 & 1 & 0 \\ s & 0 & c \end{bmatrix} \begin{bmatrix} c & -s & 0 \\ s & c & 0 \\ 0 & 0 & 1 \end{bmatrix}^d .$$

It is easy to declare that we can always just pick a lowest-degree representative, but as the matrix representive on the right of (1) shows, even in an example as simple as this one, it is rarely clear that the given representative can be further simplified. Of course with a more careful treatment we may circumvent this issue and still define degree rigorously for a quotient model. However we do not wish to distract the reader with such tangential matters. The small price we pay is that we could not speak of "linear, quadratic, cubic, quartic" programming or "degree" when discussing quotient models.

Just as we need a model for an abstract manifold to discuss manifold optimization, we need a model of complexity to discuss NP-hardness — Cook–Karp–Levin, Blum–Shub–Smale, Valiant, or yet something else. We will use that of Cook–Karp–Levin, more commonly known as the Turing *bit model* in optimization and widely adopted by optimization theorists; see [15, Section 1.1] and [42, Chapter 2]. But on occasion we will instead use the *oracle model* [15, Section 1.2], also called black-box model in [42, Chapter 6]. This is a dichotomy that we will need to highlight:

(a) In polynomial optimization, the input consists of the polynomial's coefficients, assumed rational, and thus bit complexity is well-defined.

(b) In convex optimization, the input is a convex function; and in general there is no way to quantify how many bits it takes to describe an arbitrary convex function. The standard approach around this conundrum is to use oracle complexity [32, 34].

Unless specified otherwise all of our NP-hardness results will be in terms of the more precise bit complexity (a). In a handful of less important corollaries when a bit complexity version is out of reach we resort to the oracle complexity model (b) — these will be clearly specified.

We will also reproduce [7, Definition 2.5] for FPTAS below for easy reference.

Definition 3.1 (Fully polynomial-time approximation scheme)**.** With respect to the maximization problem over $\mathcal{M}$ and the function class $\mathscr{F}$, an algorithm $\mathscr{A}$ is called a fully polynomial-time approximation scheme or FPTAS if:

(i) For any instance $f \in \mathscr{F}$ and any $\varepsilon > 0$, $\mathscr{A}$ takes the defining parameters of f (e.g., coefficients of f when f is a polynomial), ε, and $\mathcal{M}$ as input and computes an $x_\varepsilon \in \mathcal{M}$ such that $f(x_\varepsilon)$ is a $(1-\varepsilon)$-approximation of f, i.e.,

$$f_{\max} - f(x_\varepsilon) \le \varepsilon(f_{\max} - f_{\min}).$$

(ii) The number of operations required for the computation of x_ε is bounded by a polynomial in the problem size, and $1/\varepsilon$.

If $\mathscr{A}$ is an FPTAS that relies on an oracle of f for the value of f at some point, then we call it an oracle-FPTAS.

4. NP-hardness of QP over $\mathrm{Gr}(k,n)$ without fixing k and n

In this and the next two sections, we will use the projection model $\mathrm{Gr}_\pi(k,n)$ in Table 2 to model our Grassmannian. For notational simplicity we drop the subscript π and just write

$$\mathrm{Gr}(k,n) = \{P \in \mathbb{R}^{n\times n} : P^2 = P = P^\mathsf{T},\ \mathrm{tr}(P) = k\}. \tag{2}$$

Given that the Grassmannian parameterizes k-planes in n-space, one might expect to find a relation with the combinatorial problem of choosing a k-element set from an n-element set. This intuition pans out — there is a natural correspondence between the two.

Let $G = (V,E)$ be an n-vertex undirected graph with $V = \{1,\dots,n\}$. We denote an edge as an ordered pair (i,j). For an undirected graph, this means that if $(i,j) \in E$, then $(j,i) \in E$, and so a sum over E sums each edge twice. We do not allow self loop so if $(i,j) \in E$, then $i \neq j$.

The clique decision problem asks if a clique of size k exists in G. The problem is one of Karp's original 21 NP-complete problems [20]. Let $e_1,\dots,e_n \in \mathbb{R}^n$ be its standard basis. Consider the following maximization problem

$$\max_{P\in\mathrm{Gr}(k,n)} f(P) \coloneqq \max_{P\in\mathrm{Gr}(k,n)} \Biggl[\sum_{(i,j)\in E} e_i^\mathsf{T} P e_i e_j^\mathsf{T} P e_j + \sum_{i\in V} e_i^\mathsf{T} P e_i e_i^\mathsf{T} P e_i\Biggr]. \tag{3}$$

As a manifold optimization problem over $\mathrm{Gr}(k,n)$, the problem is unconstrained, and clearly quadratic, in fact homogeneous of degree two as $f(cP) = c^2 f(P)$ for any $c \in \mathbb{R}$.

Proposition 4.1 (Clique decision as Grassmannian QP)**.** *Any n-vertex graph G contains a k-clique if and only if $\max_{P\in\mathrm{Gr}(k,n)} f(P) = k^2$.*

Proof. Any projection matrix $P = (p_{ij}) \in \mathrm{Gr}(k,n)$ corresponds to the unique k-dimensional subspace $\mathrm{im}(P) \subseteq \mathbb{R}^n$. Observe that $p_{ii} = e_i^\mathsf{T} P e_i = 1$ if and only if $e_i \in \mathrm{im}(P)$. If G contains a k-clique, we may choose P to be the diagonal matrix D with $d_{ii} = 1$ when i is a vertex in the k-clique and zero everywhere else, then $f(D) = k^2$. As $P \in \mathrm{Gr}(k,n)$ is automatically positive semidefinite, we always have $p_{ii} \geq 0$ for all $i = 1,\dots,n$. This implies

$$f(P) = \sum_{(i,j)\in E} p_{ii}p_{jj} + \sum_{i\in V} p_{ii}^2 \leq \sum_{i\neq j} p_{ii}p_{jj} + \sum_{i=1}^n p_{ii}^2 = \mathrm{tr}(P)^2 = k^2.$$

So when there is a k-clique, we have $\max_{P\in\mathrm{Gr}(k,n)} f(P) = k^2$. For any $P \in \mathrm{Gr}(k,n)$, we may also write

$$f(P) = \sum_{i\neq j} p_{ii}p_{jj} + \sum_{i\in V} p_{ii}^2 - \sum_{(i,j)\notin E} p_{ii}p_{jj} = k^2 - \sum_{(i,j)\notin E} p_{ii}p_{jj}.$$

The diagonal entries satisfy $p_{11} + \cdots + p_{nn} = \mathrm{tr}(P) = k$ and $0 \leq p_{11},\dots,p_{nn} \leq 1$. We may assume that $1 \geq p_{11} \geq p_{22} \geq \cdots \geq p_{nn} \geq 0$. Then

$$k = \sum_{i=1}^n p_{ii} \leq (k-1) + (n-k+1)p_{kk},$$

so $p_{kk} \geq 1/(n-k-1)$. If G contains no k-clique, then in particular the vertices $1,\dots,k$ do not form a clique. So $(i_0,j_0) \notin E$ for some $i_0, j_0 \in \{1,\dots,k\}$. Hence $p_{i_0i_0}, p_{j_0j_0} \geq p_{kk}$,

$$\sum_{(i,j)\notin E} p_{ii}p_{jj} \geq p_{kk}^2 \geq \frac{1}{(n-k-1)^2},$$

and

$$f(P) \leq k^2 - \frac{1}{(n-k-1)^2},$$

as required. □

Applying Definition 3.1 to the current scenario, the input data for $\mathcal{M} = \operatorname{Gr}(k, n)$ is just k and n; and the input data for an instance $f \in \mathscr{F}$, the function class in (3), is just a graph $G = (V, E)$. We may deduce the following hardness result from Proposition 4.1.

Theorem 4.2 (Grassmannian quadratic programming is NP-hard I)**.** *Unless* P = NP*, there is no* FPTAS *that is polynomial in n and k for maximizing quadratic polynomials over* $\operatorname{Gr}(k, n)$*.*

Proof. If there is a k-clique in G, then $f_{\max} = k^2$ by Proposition 4.1. Take

$$\varepsilon = \frac{1}{2k^2(n-k-1)^2},$$

so that $1/\varepsilon$ is polynomial in k, n. Then an FPTAS gives a $P_\varepsilon \in \operatorname{Gr}(k, n)$ such that

$$f(P_\varepsilon) \ge f_{\max} - \varepsilon(f_{\max} - f_{\min}) \ge k^2 - \frac{1}{2k^2(n-k-1)^2}k^2 = k^2 - \frac{1}{2(n-k-1)^2}.$$

If there is no k-clique in G, then an FPTAS algorithm gives a $P_\varepsilon \in \operatorname{Gr}(k, n)$ such that

$$f(P_\varepsilon) \le f_{\max} \le k^2 - \frac{1}{(n-k-1)^2}.$$

In other words, we can decide existence or non-existence of a k-clique in G in polynomial time if there were an FPTAS for the maximization problem. □

5. NP-hardness of QP over $\operatorname{Gr}(k, n)$ with fixed k

We remind the reader that NP-hardness is a notion of *asymptotic* time complexity. In the last section, we assume that k and n both grow to infinity. In this section we will prove a stronger variant allowing $k \in \mathbb{N}$ to be arbitrary but fixed and only n grows to infinity.

We let $G = (V, E)$ be an n-vertex undirected graph following conventions set in the last section. For any $n, k \in \mathbb{N}$ with $k \le n$, we define the set

$$\Delta_{k,n} \coloneqq \{x \in \mathbb{R}^n : 0 \le x_i \le 1,\ i = 1, \dots, n,\ x_1 + \dots + x_n = k\},$$

and the function $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) \coloneqq \sum_{(i,j) \in E} x_i x_j.$$

Recall that $S \subseteq V$ is a clique if $(i, j) \in E$ for all $i, j \in S$. We will show that the *clique number*

$$\omega(G) \coloneqq \max\{|S| : S \subseteq V \text{ is a clique}\}$$

of G may be determined by maximizing f over $\Delta_{k,n}$. Indeed, for the special case $k = 1$,

$$\Delta_{1,n} = \{x \in \mathbb{R}^n : x_1 + \dots + x_n = 1,\ x_i \ge 0,\ i = 1, \dots, n\}$$

is just the unit simplex and we have the celebrated result of Motzkin and Straus [30] that

$$\max_{x \in \Delta_{1,n}} f(x) = 1 - \frac{1}{\omega(G)}. \tag{4}$$

Note that because of our convention of summing over each undirected edge twice, our expression is slightly neater, lacking the factor 1/2 on the right found in [30]. We will extend this to $\Delta_{k,n}$.

Proposition 5.1 (Generalized Motzkin–Straus)**.** *Let $k \le n$ be positive integers and G be an n-vertex undirected graph. If $\omega(G) \ge k$, then*

$$\max_{x \in \Delta_{k,n}} f(x) = k^2\left(1 - \frac{1}{\omega(G)}\right).$$

Proof. Suppose $S = \{1, \dots, \omega(G)\} \subseteq V$ is a largest clique. Then setting $x_* \in \Delta_{k,n}$ with coordinates

$$x_1^* = \cdots = x_{\omega(G)}^* = \frac{k}{\omega(G)}, \quad x_{\omega(G)+1}^* = \cdots = x_n^* = 0$$

gives us the value

$$f(x_*) = k^2\biggl(1 - \frac{1}{\omega(G)}\biggr).$$

So we only need to show that

$$\max_{x \in \Delta_{k,n}} f(x) \le k^2\biggl(1 - \frac{1}{\omega(G)}\biggr).$$

By a change-of-variable $x = ky$, the maximization problem becomes

$$\max_{y \in \Theta_{k,n}} k^2 f(y).$$

where

$$\Theta_{k,n} \coloneqq \biggl\{ y \in \mathbb{R}^n : y_1 + \cdots + y_n = 1,\ 0 \le y_i \le \frac{1}{k},\ i = 1, \dots, n \biggr\} \subseteq \Delta_{1,n}.$$

So by the original Motzkin–Straus Theorem (4),

$$\max_{x \in \Delta_{k,n}} f(x) = k^2 \max_{y \in \Theta_{k,n}} f(y) \le k^2 \max_{y \in \Delta_{1,n}} f(y) \le k^2\biggl(1 - \frac{1}{\omega(G)}\biggr),$$

as required. □

While we will not need this, but the statement in Proposition 5.1 is really "if and only if." We may show that if $\omega(G) < k$, then

$$\max_{x \in \Delta_{k,n}} f(x) < k^2\biggl(1 - \frac{1}{\omega(G)}\biggr).$$

In this sense the result in Proposition 5.1 is sharp.

Proposition 5.2 (Clique number as Grassmannian QP)**.** *Let $k \le n$ be positive integers and G be an n-vertex undirected graph. If $\omega(G) \ge k$, then*

$$\max_{P \in \operatorname{Gr}(k,n)} \sum_{(i,j) \in E} p_{ii} p_{jj} = k^2\biggl(1 - \frac{1}{\omega(G)}\biggr).$$

Proof. Here $\operatorname{Gr}(k,n)$ is as in (2). By the Schur–Horn Theorem [18], the diagonal map $\operatorname{diag} : \mathbb{R}^{n \times n} \to \mathbb{R}^n$, when restricted to $\operatorname{Gr}(k,n)$, gives a surjection

$$\operatorname{diag} : \operatorname{Gr}(k,n) \to \Delta_{k,n}, \quad P \mapsto \operatorname{diag}(P).$$

In other words, with a relabeling of variables $x_i = p_{ii}$, $i = 1, \dots, n$, we get

$$\max_{P \in \operatorname{Gr}(k,n)} \sum_{(i,j) \in E} p_{ii} p_{jj} = \max_{x \in \Delta_{k,n}} \sum_{(i,j) \in E} x_i x_j.$$

So the required result follows from Proposition 5.1. As in the previous section, the objective function may be written $\sum_{(i,j) \in E} e_i^{\mathsf{T}} P e_i e_j^{\mathsf{T}} P e_j$, which is clearly quadratic in P. □

Note that our next proof will rely on the fact that for any fixed k, the time complexity of the k-clique problem is polynomial in n. Even an exhaustive search through all subgraphs with k or fewer vertices has time $\binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k}$, a degree-$k$ polynomial in n.

Theorem 5.3 (Grassmannian quadratic programming is NP-hard II)**.** *Let $k \in \mathbb{N}$ be fixed. Unless* P = NP*, there is no* FPTAS *that is polynomial in n for maximizing quadratic polynomials over* $\operatorname{Gr}(k,n)$*.*

Proof. Given any n-vertex graph $G = (V, E)$ as input, we may test all subgraphs with k or fewer vertices and check if $\omega(G) < k$ in time polynomial in n. If $\omega(G) < k$, this also yields the exact value of $\omega(G)$. If $\omega(G) \ge k$, then we find an ε-approximate solution to

$$\max_{P \in \operatorname{Gr}(k,n)} \sum_{(i,j)\in E} p_{ii} p_{jj}$$

using an FPTAS, assuming one exists, where

$$\varepsilon = \frac{1}{2k^2 n^2}.$$

The maximum only takes finitely many different discrete values depending on what $\omega(G)$ is, and these values have relative gaps of at least ε. So if the maximum value can be approximated to ε-accuracy, then by Proposition 5.1 we get the exact value of $\omega(G)$. Hence unless P = NP, there is no FPTAS for maximizing quadratic polynomials over $\operatorname{Gr}(k, n)$. □

Since orthogonal projection matrices are always positive semidefinite, we have

$$\operatorname{Gr}(1, n) = \{P \in \mathbb{S}^n_+ : P^2 = P,\ \operatorname{tr}(P) = 1\},$$

i.e., it is exactly the set of density matrices of pure states. Its convex hull is then the set of all *density matrices* of mixed states [36, Theorem 2.5],

$$\mathscr{D}_n \coloneqq \{X \in \mathbb{S}^n_+ : \operatorname{tr}(X) = 1\} = \operatorname{conv}\operatorname{Gr}(1, n).$$

From this we obtained the following slightly unexpected payoff.

Corollary 5.4 (Quadratic programming over density matrices is NP-hard)**.** *Unless* P = NP, *there is no* FPTAS *that is polynomial in n for maximizing quadratic polynomials over* $\mathscr{D}_n$.

Proof. Following the proof of Proposition 5.2, the diagonal map $\operatorname{diag} : \mathbb{R}^{n\times n} \to \mathbb{R}^n$, when restricted to $\mathscr{D}_n$, gives a surjection

$$\operatorname{diag} : \mathscr{D}_n \to \Delta_{1,n}, \quad X \mapsto \operatorname{diag}(X).$$

Relabeling $x_i = x_{ii}$, $i = 1, \dots, n$, we get

$$\max_{X\in\mathscr{D}_n} \sum_{(i,j)\in E} x_{ii} x_{jj} = \max_{x\in\Delta_{1,n}} \sum_{(i,j)\in E} x_i x_j = 1 - \frac{1}{\omega(G)}$$

by (4). Now it remains to repeat, with $k = 1$, the same argument in the second paragraph of the proof of Theorem 5.3. □

Observe that we would not have been able to obtain this using the results in Section 4, which do not allow k to be fixed.

6. NP-hardness for other models of the Grassmannian

We give a list of explicit, polynomial-time change-of-coordinates formulas for transformation between various models of the Grassmannian, from which it will follow that our NP-hardness results in Sections 4 and 5 apply to all models of the Grassmannian in Tables 1 and 2.

Proposition 6.1. *There exist maps* $\varphi_1, \dots, \varphi_5$ *in the diagram below*

$$\begin{tikzcd} \operatorname{O}(n)/(\operatorname{O}(n-k)\times\operatorname{O}(k)) \arrow[r, "\varphi_1"] \arrow[d, "\varphi_5"'] & \operatorname{V}(k,n)/\operatorname{O}(k) \arrow[r, "\varphi_2"] \arrow[d, "\varphi_4"'] & \operatorname{Gr}_\pi(k,n) \arrow[r, "\varphi_3"] & \operatorname{Gr}_{a,b}(k,n) \\ \operatorname{GL}(n)/\operatorname{P}(k,n) & \operatorname{St}(k,n)/\operatorname{GL}(k) & & \end{tikzcd} \tag{5}$$

that are diffeomorphisms of smooth manifolds, and polynomial-time computable.

Proof. We will give each of these maps and its inverse explicitly. As there are multiple quotient manifolds involved in (5), we distinguish them by labeling with self-explanatory identifiers in subscript.

The map $\varphi_1 : \mathrm{O}(n)/(\mathrm{O}(n-k)\times\mathrm{O}(k)) \to \mathrm{V}(k,n)/\mathrm{O}(k)$ and its inverse are given by

$$\varphi_1([\![Q]\!]_{\mathrm{O}}) = \left[\!\!\left[Q\begin{bmatrix} I_k \\ 0\end{bmatrix}\right]\!\!\right]_{\mathrm{V}}, \quad \varphi_1^{-1}([\![Y]\!]_{\mathrm{V}}) = [\![Q_Y]\!]_{\mathrm{O}}.$$

Here $Q_Y \in \mathrm{O}(n)$ is the eigenvector matrix in the eigenvalue decomposition $YY^\mathsf{T} = Q_Y\left[\begin{smallmatrix} I_k & 0 \\ 0 & 0\end{smallmatrix}\right]Q_Y^\mathsf{T}$. Note that multiplying an $n\times n$ matrix on the right by $\left[\begin{smallmatrix} I_k \\ 0\end{smallmatrix}\right] \in \mathbb{R}^{n\times k}$ gives an $n\times k$ matrix comprising the first k columns. We check that

$$\begin{aligned}\varphi_1^{-1}\circ\varphi_1([\![Q]\!]_{\mathrm{O}}) &= \varphi_1^{-1}\left(\left[\!\!\left[Q\begin{bmatrix} I_k \\ 0\end{bmatrix}\right]\!\!\right]_{\mathrm{V}}\right) = [\![Q]\!]_{\mathrm{O}},\\ \varphi_1\circ\varphi_1^{-1}([\![Y]\!]_{\mathrm{V}}) &= \varphi_1([\![Q_Y]\!]_{\mathrm{O}}) = \left[\!\!\left[Q_Y\begin{bmatrix} I_k \\ 0\end{bmatrix}\right]\!\!\right]_{\mathrm{V}} = [\![Y]\!]_V.\end{aligned}$$

Here the last equality follows from the observation that $ZZ^\mathsf{T} = YY^\mathsf{T}$ for $Z, Y \in \mathrm{V}(k,n)$ if and only if $Z = YP$ for some $P \in \mathrm{O}(k)$.

The map $\varphi_2 : \mathrm{V}(k,n)/\mathrm{O}(k) \to \mathrm{Gr}_\pi(k,n)$ and its inverse are given by

$$\varphi_2([\![Y]\!]_{\mathrm{V}}) = YY^\mathsf{T}, \quad \varphi_2^{-1}(P) = \left[\!\!\left[Q_P\begin{bmatrix} I_k \\ 0\end{bmatrix}\right]\!\!\right]_{\mathrm{V}}.$$

Here $Q_P \in \mathrm{O}(n)$ is the eigenvector matrix in the eigenvalue decomposition $P = Q_P\left[\begin{smallmatrix} I_k & 0 \\ 0 & 0\end{smallmatrix}\right]Q_P^\mathsf{T}$. We check that

$$\varphi_2^{-1}\circ\varphi_2([\![Y]\!]_{\mathrm{V}}) = \varphi_2^{-1}(YY^\mathsf{T}) = [\![Y]\!]_{\mathrm{V}},$$

since the column vectors of Y are exactly the nonzero eigenvectors of the orthogonal projection matrix YY^T. We also check that

$$\varphi_2\circ\varphi_2^{-1}(P) = \varphi_2\left(\left[\!\!\left[Q_P\begin{bmatrix} I_k \\ 0\end{bmatrix}\right]\!\!\right]_{\mathrm{V}}\right) = Q_P\begin{bmatrix} I_k & 0 \\ 0 & 0\end{bmatrix}Q_P^\mathsf{T} = P.$$

The map $\varphi_3 : \mathrm{Gr}_\pi(k,n) \to \mathrm{Gr}_{a,b}(k,n)$ and its inverse are given by

$$\varphi_3(P) = (a-b)P + bI_n, \quad \varphi_3^{-1}(W) = \frac{1}{a-b}(W - bI_n),$$

which is obvious by inspection.

The map $\varphi_4 : \mathrm{V}(k,n)/\mathrm{O}(k) \to \mathrm{St}(k,n)/\mathrm{GL}(k)$ and its inverse are given by

$$\varphi_4([\![Y]\!]_{\mathrm{V}}) = [\![Y]\!]_{\mathrm{St}}, \quad \varphi_4^{-1}([\![S]\!]_{\mathrm{St}}) = [\![Y_S]\!]_{\mathrm{V}}.$$

Here $Y_S \in \mathrm{V}(k,n)$ is the unique orthogonal factor in the condensed QR decomposition $S = Y_S R_S$ such that $R_S \in \mathbb{R}^{k\times k}$ is upper triangular with positive diagonal. We check that

$$\begin{aligned}\varphi_4^{-1}\circ\varphi_4([\![Y]\!]_{\mathrm{V}}) &= \varphi_4^{-1}([\![Y]\!]_{\mathrm{St}}) = [\![Y]\!]_{\mathrm{V}},\\ \varphi_4\circ\varphi_4^{-1}([\![S]\!]_{\mathrm{St}}) &= \varphi_4([\![Y_S]\!]_{\mathrm{V}}) = [\![Y_S]\!]_{\mathrm{St}} = [\![Y_S R_S]\!]_{\mathrm{St}} = [\![S]\!]_{\mathrm{St}}.\end{aligned}$$

The map $\varphi_5 : \mathrm{O}(n)/(\mathrm{O}(n-k)\times\mathrm{O}(k)) \to \mathrm{GL}(n)/\mathrm{P}(k,n)$ and its inverse are given by

$$\varphi_5([\![Q]\!]_{\mathrm{O}}) = [\![Q]\!]_{\mathrm{GL}}, \quad \varphi_5^{-1}([\![X]\!]_{\mathrm{GL}}) = [\![Q_X]\!]_{\mathrm{O}}.$$

Here $Q_X \in \mathrm{O}(n)$ is the unique orthogonal matrix in the full QR-decomposition $X = Q_X R_X$ where $R_X \in \mathbb{R}^{n\times n}$ is upper triangular with positive diagonal. Since $R_X \in \mathrm{P}(k,n)$, we have

$$\begin{aligned}\varphi_5^{-1}\circ\varphi_5([\![Q]\!]_{\mathrm{O}}) &= \varphi_5^{-1}([\![Q]\!]_{\mathrm{GL}}) = [\![Q]\!]_{\mathrm{O}},\\ \varphi_5\circ\varphi_5^{-1}([\![X]\!]_{\mathrm{GL}}) &= \varphi_1([\![Q_X]\!]_{\mathrm{O}}) = [\![Q_X]\!]_{\mathrm{GL}} = [\![Q_X R_X]\!]_{\mathrm{GL}} = [\![X]\!]_{\mathrm{GL}}.\end{aligned}$$

Each φ_i is smooth and its differential $d\varphi_i$ is an isomorphism at every point. This is clear for φ_3 as it is an invertible linear map. This holds for φ_2 as it is induced by the smooth map $\overline{\varphi}_2 : \mathrm{V}(k,n) \to \mathrm{Gr}_\pi(k,n)$, $Y \mapsto YY^\mathsf{T}$; and $d_{[\![Y]\!]}\varphi_2$ is an isomorphism as $d_Y\overline{\varphi}_2$ is surjective and its kernel is isomorphic to the Lie algebra $\mathfrak{o}(k)$ of $k \times k$ skew symmetric matrices. The arguments for φ_3, φ_4, φ_5 are similar to that for φ_2. Hence $\varphi_1, \dots, \varphi_5$ are all diffeomorphisms since they are also bijective [25, Theorem 4.14].

Finally, since these diffeomorphisms $\varphi_1, \dots, \varphi_5$ and their inverses can be computed with matrix-matrix products, QR decomposition, and symmetric eigenvalue decomposition, all of which have $\mathcal{O}(n^\omega)$ time complexity, with ω the exponent of matrix multiplication — see [38] and [1, Remark 16.26]. We conclude that these diffeomorphisms are all polynomial-time computable. □

By Proposition 6.1, every arrow in (5) may be reversed, allowing one to get from any model to any other model in (5) by simply composing and inverting maps, with the resulting map remaining a polynomial-time computable diffeomorphism. But the reader is reminded of the dichotomy we highlighted on p. 4 — having a polynomial-time computable diffeomorphism does not translate directly to polynomial-time reduction under the bit complexity model.

Take $\varphi_2 : \mathrm{V}(k,n)/\,\mathrm{O}(k) \to \mathrm{Gr}_\pi(k,n)$ for illustration. Saying that it is a polynomial-time computable diffeomorphism means that if we can evaluate $f(P)$ for some $f : \mathrm{Gr}_\pi(k,n) \to \mathbb{R}$ and any $P \in \mathrm{Gr}_\pi(k,n)$ in a certain time complexity, then we can evaluate $f \circ \varphi_2([\![Y]\!])$ for any $[\![Y]\!] \in \mathrm{V}(k,n)/\,\mathrm{O}(k)$ with an increase in time complexity of at most a polynomial factor. It does not mean that we have a polynomial-time algorithm to transform the defining parameters of f into the defining parameters of $f \circ \varphi_2$.

Corollary 6.2. *Let $f : \mathrm{Gr}_\pi(k,n) \to \mathbb{R}$ be a polynomial function of bounded degree. Then f can be transformed in polynomial-time into:*

(a) *a polynomial function on $\mathrm{Gr}_{a,b}(k,n)$ of the same degree as f and whose coefficients are affine combinations of the coefficients of f;*

(b) *a real-valued function on $\mathrm{O}(n)/(\mathrm{O}(n-k) \times \mathrm{O}(k))$ or $\mathrm{V}(k,n)/\,\mathrm{O}(k)$ defined by parameters given by the same coefficients of f but repeated with some multiplicities.*

Proof. From the proof of Proposition 6.1, $\varphi_3^{-1} : \mathrm{Gr}_{a,b}(k,n) \to \mathrm{Gr}_\pi(k,n)$ is an invertible affine map. So $f \circ \varphi_3^{-1}$ is a polynomial function of the same degree as f whose coefficients are given by affine combinations of those in f.

For the map $\varphi_1 : \mathrm{O}(n)/(\mathrm{O}(n-k) \times \mathrm{O}(k)) \to \mathrm{V}(k,n)/\,\mathrm{O}(k)$, observe that no matter which representative $Q \in \mathrm{O}(n)$ we pick for $[\![Q]\!]_\mathrm{O}$, the matrix $Q\left[\begin{smallmatrix} I_k \\ 0 \end{smallmatrix}\right] \in \mathrm{V}(k,n)$ will always be a valid representative for $\varphi_1([\![Q]\!]_\mathrm{O})$. For the map $\varphi_2 : \mathrm{V}(k,n)/\,\mathrm{O}(k) \to \mathrm{Gr}_\pi(k,n)$, observe that no matter which representative $Y \in \mathrm{V}(k,n)$ we pick for $[\![Y]\!]_\mathrm{V}$, it will always be mapped to the same matrix $YY^\mathsf{T} \in \mathrm{Gr}_\pi(k,n)$. So these functions are well-defined, i.e., they give the same value irrespective of the representative picked.

The value of $f \circ \varphi_2$ on $[\![Y]\!]_\mathrm{V} \in \mathrm{V}(k,n)/\,\mathrm{O}(k)$ is given by $f(YY^\mathsf{T})$ where $Y \in \mathrm{V}(k,n)$ is any representative of $[\![Y]\!]_\mathrm{V}$. The value of $f \circ \varphi_2 \circ \varphi_1$ on $[\![Q]\!]_\mathrm{O} \in \mathrm{O}(n)/(\mathrm{O}(n-k) \times \mathrm{O}(k))$ is given by $f\bigl(Q\left[\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right]Q^\mathsf{T}\bigr)$ where $Q \in \mathrm{O}(n)$ is any representative of $[\![Q]\!]_\mathrm{O}$. In other words the parameters defining the map $f \circ \varphi_2$ or $f \circ \varphi_2 \circ \varphi_1$ are just copies of the coefficients of f. □

To explain the statement "repeated with some multiplicities" in Corollary 6.2(b) more precisely, we provide a concrete example.

Example 6.3 (Quadratic case)**.** Suppose $f : \mathrm{Gr}(k,n) \to \mathbb{R}$ is a quadratic function, i.e.,

$$f(P) = \sum_{i,j,r,s=1}^{n} a_{ijrs} p_{ij} p_{rs} + \sum_{i,j=1}^{n} b_{ij} p_{ij} + c. \tag{6}$$

We assume $c = 0$. Then on $\mathrm{V}(k,n)/\mathrm{O}(k)$,

$$\begin{aligned}(7)\quad f\circ\varphi_2(\llbracket Y\rrbracket_{\mathrm{V}}) &= f(YY^{\mathsf{T}})\\ &= \sum_{i,j,r,s=1}^{n} a_{ijrs}(y_{i1}y_{j1}+\dots+y_{ik}y_{jk})(y_{r1}y_{s1}+\dots+y_{rk}y_{sk}) + \sum_{i,j=1}^{n} b_{ij}(y_{i1}y_{j1}+\dots+y_{ik}y_{jk})\end{aligned}$$

where $Y \in \mathrm{V}(k,n)$ is any representative of $\llbracket Y\rrbracket_{\mathrm{V}} \in \mathrm{V}(k,n)/\mathrm{O}(k)$. To be even more concrete, suppose $k = 2$, then the (i,j,r,s)th term in the first sum takes the form

$$a_{ijrs}y_{i1}y_{j1}y_{r1}y_{s1} + a_{ijrs}y_{i2}y_{j2}y_{r1}y_{s1} + a_{ijrs}y_{i1}y_{j1}y_{r2}y_{s2} + a_{ijrs}y_{i2}y_{j2}y_{r2}y_{s2}$$

and the (i,j)th term in the second sum takes the form

$$b_{ij}y_{i1}y_{j1} + b_{ij}y_{i2}y_{j2}.$$

In general, the coefficient a_{ijrs} in f will be repeated k^2 times in the defining parameters of $f\circ\varphi_2$ and the coefficient b_{ijrs} will be repeated k times. More generally, a degree-d term in f will see its coefficient repeated k^d times as defining parameters of $f\circ\varphi_2$. This discussion applies verbatim to $f\circ\varphi_2\circ\varphi_1$. □

We now state our main result of this section.

Theorem 6.4 (Grassmannian optimization is NP-hard I)**.** *Let $k \in \mathbb{N}$ be fixed. Unless* P = NP*, there is no* FPTAS *that is polynomial in n for:*

(i) *maximizing quadratic polynomials over $\mathrm{Gr}_{a,b}(k,n)$;*
(ii) *maximizing real-valued functions over $\mathrm{O}(n)/(\mathrm{O}(n-k)\times\mathrm{O}(k))$ or $\mathrm{V}(k,n)/\mathrm{O}(k)$.*

Proof. If there were an FPTAS for maximizing real-valued functions on $\mathrm{V}(k,n)/\mathrm{O}(k)$, we have an FPTAS for maximizing any quadratic function $f : \mathrm{Gr}_\pi(k,n) \to \mathbb{R}$ by maximizing

$$f\circ\varphi_2 : \mathrm{V}(k,n)/\mathrm{O}(k) \to \mathbb{R},$$

contradicting Theorem 5.3. The same argument applies to $f\circ\varphi_3^{-1}$ and $f\circ\varphi_2\circ\varphi_1$ for the other two models. □

Note that Theorem 6.4 is in terms of standard bit complexity. But for the next result we will need to use the oracle complexity mentioned on p. 4.

Corollary 6.5 (Grassmannian optimization is NP-hard II)**.** *Let $k \in \mathbb{N}$ be fixed. Unless* P = NP*, there is no oracle-*FPTAS *that is polynomial in n for maximizing real-valued functions over $\mathrm{St}(k,n)/\mathrm{GL}(k)$ or $\mathrm{GL}(n)/\mathrm{P}(k,n)$.*

Proof. By Proposition 6.1,

$$\varphi_2 : \mathrm{V}(k,n)/\mathrm{O}(k) \to \mathrm{Gr}_\pi(k,n), \quad \varphi_4^{-1} : \mathrm{St}(k,n)/\mathrm{GL}(k) \to \mathrm{V}(k,n)/\mathrm{O}(k)$$

are both polynomial-time computable diffeomorphisms. If there were an FPTAS for maximizing real-valued functions on $\mathrm{St}(k,n)/\mathrm{GL}(k)$, we have an FPTAS for maximizing any quadratic function $f : \mathrm{Gr}_\pi(k,n) \to \mathbb{R}$ by querying

$$f\circ\varphi_2\circ\varphi_4^{-1} : \mathrm{St}(k,n)/\mathrm{GL}(k) \to \mathbb{R}$$

as an oracle, contradicting Theorem 5.3. The same argument also applies to

$$f\circ\varphi_2\circ\varphi_1\circ\varphi_5^{-1} : \mathrm{GL}(n)/\mathrm{P}(k,n) \to \mathbb{R}.$$ □

The proof of Corollary 6.5 superficially resembles that of Corollary 6.2; the difference being that φ_1 and φ_2 are simple polynomial maps, whereas φ_4^{-1} and φ_5^{-1} involve QR decompositions.[2] While

[2] This is inevitable. The inverse of any isometry from $\mathrm{St}(k,n)/\mathrm{GL}(k)$ or $\mathrm{GL}(n)/\mathrm{P}(k,n)$ to $\mathrm{Gr}_\pi(k,n)$ must necessarily involve the QR-decomposition [13].

polynomial-time algorithms for computing QR decomposition of a full-rank matrix abound (this is why φ_4^{-1} and φ_5^{-1} can be evaluated on any specific input in polynomial-time), they invariably involve extracting square roots and branching. There is no known rational map that can replace an algorithm for QR decomposition; should one exists, it is expected to be exceedingly complicated so that composing it with a polynomial f would likely result in the defining parameters blowing up to exponentially high degrees.

7. NP-hardness for Stiefel manifolds and orthogonal group

As we mentioned in Section 1, unconstrained *quadratic* programming on $\mathrm{V}(1,n) = \mathrm{S}^{n-1}$ is polynomial-time [42, Section 4.3]. On the other hand, it is an immediate corollary of Proposition 5.2 that for any fixed $k \le n$, unconstrained *quartic* programming on $\mathrm{V}(k,n)$ is NP-hard. In fact we may also deduce the same for $\mathrm{O}(n)$, though not by simply setting $\mathrm{O}(n) = \mathrm{V}(n,n)$ as k is fixed:

Corollary 7.1 (Stiefel and orthogonal quartic programming is NP-hard)**.** *Let $k \in \mathbb{N}$ be fixed. Unless* P = NP*, there is no* FPTAS *that is polynomial in n for maximizing quartic polynomials over* $\mathrm{V}(k,n)$ *and* $\mathrm{O}(n)$*.*

Proof. Let $\mathrm{Gr}(k,n)$ be the projection model (2). We define the two maps

$$\mathrm{O}(n) \xrightarrow{\pi_1} \mathrm{V}(k,n) \xrightarrow{\pi_2} \mathrm{Gr}(k,n), \qquad \pi_1(Q) := Q\begin{bmatrix} I_k \\ 0 \end{bmatrix}, \qquad \pi_2(Y) := YY^{\mathsf{T}},$$

noting that right multiplication by $\left[\begin{smallmatrix} I_k \\ 0 \end{smallmatrix}\right] \in \mathbb{R}^{n\times k}$ gives a matrix in $\mathrm{V}(k,n)$ comprising the first k columns of Q. Let $f : \mathrm{Gr}(k,n) \to \mathbb{R}$ be a quadratic function as in (6). Then the functions $f \circ \pi_2 : \mathrm{V}(k,n) \to \mathbb{R}$ and $f \circ \pi_2 \circ \pi_1 : \mathrm{O}(n) \to \mathbb{R}$ are given by

$$f \circ \pi_2(Y) = f(YY^{\mathsf{T}}), \quad f \circ \pi_2 \circ \pi_1(Q) = f\left(Q \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} Q^{\mathsf{T}}\right),$$

i.e., quartic polynomial functions in Y and Q respectively. In particular $f \circ \pi_2(Y)$ is given by the same expression in (7) and $f \circ \pi_2 \circ \pi_1$ is given by a nearly identical expression (with q_{ij} in place of y_{ij}). By Theorem 5.3, we deduce that there is no FPTAS for maximizing quartic polynomials over $\mathrm{V}(k,n)$ and $\mathrm{O}(n)$. □

It is not surprising that quartic programming is NP-hard, given that it is already NP-hard over $\mathbb{R}^n$ [31], as well as in a variety of other scenarios [19]. So it remains to find out if unconstrained *cubic* programming on $\mathrm{V}(k,n)$ is NP-hard. We will show that it is but this cannot be easily deduced from any of our results up to this point — as we saw in the proof above, these give quartic programs on $\mathrm{V}(k,n)$.

We will instead use a celebrated result of Nesterov [33], which shows that $\alpha(G)$, the *stability number* of an $(n-1)$-vertex undirected graph G, may be expressed as the maximum of a homogenous cubic polynomial f over the $(n-1)$-sphere S^{n-1}. We will not need to know the actual expression for f nor the definition of stability number, just that

$$\max_{\|x\|=1} f(x) = \sqrt{1 - \frac{1}{\alpha(G)}} \tag{8}$$

and that this implies there is no FPTAS for maximizing f over S^{n-1}.

Theorem 7.2 (Stiefel and orthogonal cubic programming is NP-hard)**.** *Let $k \in \mathbb{N}$ be fixed. Unless* P = NP*, there is no* FPTAS *that is polynomial in n for maximizing cubic polynomials over* $\mathrm{V}(k,n)$ *or* $\mathrm{O}(n)$*.*

Proof. Let $e_1 \in \mathbb{R}^n$ be the first standard basis vector. Then $\pi_1 : \mathrm{V}(k,n) \to \mathrm{S}^{n-1}$, $Y \mapsto Ye_1$, takes Y to its first column vector. Let $f : \mathrm{S}^{n-1} \to \mathbb{R}$ be a polynomial function. Then $f \circ \pi_1 : \mathrm{V}(k,n) \to \mathbb{R}$ is a polynomial of the same degree. If f is Nesterov's cubic polynomial in (8), then $f \circ \pi_1$ is a cubic polynomial on $\mathrm{V}(k,n)$ with the desired property. For $\mathrm{O}(n)$, note that same argument also applies to $\pi_1 : \mathrm{O}(n) \to \mathrm{S}^{n-1}$, $Q \mapsto Qe_1$. □

Nesterov's construction also yields an independent proof of the NP-hardness result for $\mathrm{Gr}(1,n)$. We will record it below given that this is possibly the most important Grassmannian: As an abstract manifold, it is the $(n-1)$-dimensional real projective space $\mathbb{RP}^{n-1}$; and as we saw in Section 5, it may also be regarded as the set of density matrices of pure states.

Corollary 7.3 (Grassmannian quadratic programming is NP-hard III)**.** *Unless* $\mathrm{P} = \mathrm{NP}$*, there is no* FPTAS *that is polynomial in* n *for maximizing quadratic polynomials over* $\mathrm{Gr}(1,n)$*.*

Proof. Again we will use the projection model in (2) for $\mathrm{Gr}(1,n)$. Recalling that we write vectors as column vectors, let

$$\pi : \mathrm{S}^{n-1} \to \mathrm{Gr}(1,n), \quad \pi(x,y) = \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} x^{\mathsf{T}} & y \end{bmatrix} = \begin{bmatrix} xx^{\mathsf{T}} & yx \\ yx^{\mathsf{T}} & y^2 \end{bmatrix},$$

where $x \in \mathbb{R}^{n-1}$ satisfies

$$\left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\|^2 \coloneqq x_1^2 + \cdots + x_{n-1}^2 + y^2 = 1.$$

Let f be Nesterov's cubic polynomial in (8) and define the optimization problem over S^n,

$$\max_{\|x\|^2 + y^2 = 1} g(x,y) \coloneqq \max_{\|x\|^2 + y^2 = 1} f(x)y.$$

So g is a homogeneous quartic polynomial in the vector variable $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$. But observe that we may also write g in the form of a homogeneous quadratic polynomial h in the matrix variable $P = \left[\begin{smallmatrix} xx^{\mathsf{T}} & yx \\ yx^{\mathsf{T}} & y^2 \end{smallmatrix}\right]$ since π is surjective. In other words, we have $h \circ \pi = g$ and therefore

$$\begin{aligned} \max_{P \in \mathrm{Gr}(1,n)} h(P) = \max_{\|(x,y)\|=1} f(x)y = \max_{\|x\| \le 1} f(x)\sqrt{1 - \|x\|^2} &= \max_{\|x\|=1,\, 0 \le t \le 1} f(tx)\sqrt{1-t^2} \\ &= \max_{\|x\|=1,\, 0 \le t \le 1} f(x)t^3\sqrt{1-t^2} = c \max_{\|x\|=1} f(x) \end{aligned}$$

for some constant $c > 0$. This shows that there is no FPTAS for maximizing quadratic polynomials over $\mathrm{Gr}(1,n)$. □

We establish an analogue of Proposition 6.1 in order to show that the NP-hardness results above apply to the quotient model of the Stiefel manifold and to the noncompact Stiefel manifold.

Lemma 7.4. *There exist polynomial-time computable diffeomorphisms* $\psi_1 : \mathrm{O}(n)/\mathrm{O}(n-k) \to \mathrm{V}(k,n)$ *and* $\psi_2 : \mathrm{GL}(n)/\mathrm{P}_1(k,n) \to \mathrm{St}(k,n)$*.*

Proof. For the quotient manifold $\mathrm{O}(n)/\mathrm{O}(n-k)$, we regard $\mathrm{O}(n-k)$ as the subgroup

$$\left\{ \begin{bmatrix} I_k & 0 \\ 0 & Q_2 \end{bmatrix} \in \mathrm{O}(n) : Q_2 \in \mathrm{O}(n-k) \right\}.$$

With this convention, the map below is well-defined

$$\psi_1 : \mathrm{O}(n)/\mathrm{O}(n-k) \to \mathrm{V}(k,n), \quad \psi_1(\llbracket Q \rrbracket) = Q \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

and invertible. The inverse map is

$$\psi_1^{-1} : \mathrm{V}(k,n) \to \mathrm{O}(n)/\mathrm{O}(n-k), \quad \psi_1^{-1}(Y) = \llbracket Q_Y \rrbracket,$$

where Q_Y is the eigenvector matrix in the eigenvalue decomposition $YY^{\mathsf{T}} = Q_Y \left[\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right] Q_Y^{\mathsf{T}}$, and although Q_Y is not uniquely determined, its equivalence class $[\![Q_Y]\!]$ is. Hence ψ_1^{-1} is also well-defined. It is routine to verify that ψ_1^{-1} is indeed the inverse of ψ_1. The same argument in the proof of Proposition 6.1 shows that ψ_1 is a diffeomorphism. Computation of ψ_1 is simply taking a submatrix while computation of ψ_1^{-1} involves eigenvalue decomposition, but both are clearly computable in polynomial-time.

In the last paragraph, $[\![\,\cdot\,]\!] = [\![\,\cdot\,]\!]_{\mathrm{O}}$; in this paragraph $[\![\,\cdot\,]\!] = [\![\,\cdot\,]\!]_{\mathrm{GL}}$. We define

$$\psi_2 : \mathrm{GL}(n)/\,\mathrm{P}_1(k,n) \to \mathrm{St}(k,n), \quad \psi_2([\![X]\!]) = S \begin{bmatrix} I_k \\ 0 \end{bmatrix},$$

$$\psi_2^{-1} : \mathrm{St}(k,n) \to \mathrm{GL}(n)/\,\mathrm{P}_1(k,n), \quad \psi_2^{-1}(S) = [\![L_S]\!],$$

where $L_S \in \mathrm{GL}(n)$ is a unit lower-triangular matrix in an LDL^{T} decomposition

$$SS^{\mathsf{T}} = L_S D L_S^{\mathsf{T}}$$

with $D \in \mathbb{R}^{n\times n}$ a nonnegative diagonal matrix. Although L_S is not unique, but different choices differ by an element in $\mathrm{P}_1(k,n)$, and thus ψ_2^{-1} is well-defined. Again it is routine to verify that ψ_2^{-1} is the inverse of ψ_2 and that ψ_2 is a diffeomorphism. Computation of ψ_2 is simply taking a submatrix while computation of ψ_2^{-1} involves LDL^{T} decomposition, both computable in polynomial-time. □

We may now deduce that optimization over the quotient model of the compact Stiefel manifold is also NP-hard in the bit model.

Corollary 7.5 (Compact Stiefel optimization is NP-hard). *Let $k \in \mathbb{N}$ be fixed. Unless* P = NP*, there is no* FPTAS *that is polynomial in n for maximizing real-valued functions over* $\mathrm{O}(n)/\,\mathrm{O}(n-k)$*.*

Proof. Let $f : \mathrm{V}(k,n) \to \mathbb{R}$ be a polynomial function. The value of $f \circ \psi_1$ on $[\![Q]\!] \in \mathrm{O}(n)/\,\mathrm{O}(k)$ is given by $f\bigl(Q\left[\begin{smallmatrix} I_k \\ 0 \end{smallmatrix}\right]\bigr)$ where $Q \in \mathrm{O}(n)$ is any representative of $[\![Q]\!]$. In other words the parameters defining the map $f \circ \psi_1$ are the coefficients of f repeated with some multiplicities, as we saw in Example 6.3. If there were an FPTAS for maximizing real-valued functions on $\mathrm{O}(n)/\,\mathrm{O}(k)$, then we have an FPTAS for maximizing $f : \mathrm{V}(k,n) \to \mathbb{R}$, contradicting Theorem 7.2. □

Finally, we show that optimization over either the submanifold or quotient model of the noncompact Stiefel manifold, or the general linear group, is NP-hard in the oracle model.

Corollary 7.6 (Noncompact Stiefel optimization is NP-hard). *Let $k \in \mathbb{N}$ be fixed. Unless* P = NP*, there is no oracle-*FPTAS *that is polynomial in n for maximizing real-valued functions over* $\mathrm{St}(k,n)$*,* $\mathrm{GL}(n)/\,\mathrm{P}_1(k,n)$*, or* $\mathrm{GL}(n)$*.*

Proof. Let $\psi_2 : \mathrm{GL}(n)/\,\mathrm{P}_1(k,n) \to \mathrm{St}(k,n)$ be as defined in Lemma 7.4. Let $\theta_1 : \mathrm{GL}(n) \to \mathrm{GL}(n)/\,\mathrm{P}_1(k,n)$ be the quotient map taking a matrix to its coset in the quotient group. Let $\theta_2 : \mathrm{St}(k,n) \to \mathrm{V}(k,n)$ be given by $\theta_2(S) = Y_S$, where $Y_S \in \mathrm{V}(k,n)$ is the unique orthogonal factor in the condensed QR-decomposition $S = Y_S R_S$ in which $R_S \in \mathbb{R}^{k\times k}$ is upper triangular with positive diagonal. These three maps fit into a sequence:

$$\mathrm{GL}(n) \xrightarrow{\theta_1} \mathrm{GL}(n)/\,\mathrm{P}_1(k,n) \xrightarrow{\psi_2} \mathrm{St}(k,n) \xrightarrow{\theta_2} \mathrm{V}(k,n).$$

Now let $f \circ \pi_1 : \mathrm{V}(k,n) \to \mathbb{R}$ be the cubic polynomial constructed in the proof of Theorem 7.2. Then

$$\begin{gathered} f \circ \pi_1 \circ \theta_2 : \mathrm{St}(k,n) \to \mathbb{R}, \\ f \circ \pi_1 \circ \theta_2 \circ \psi_2 : \mathrm{GL}(n)/\,\mathrm{P}_1(k,n) \to \mathbb{R}, \\ f \circ \pi_1 \circ \theta_2 \circ \psi_2 \circ \theta_1 : \mathrm{GL}(n) \to \mathbb{R} \end{gathered} \tag{9}$$

are all real-valued function that can be computed in the same time complexity as $f \circ \pi_1$, up to a polynomial time factor. The required conclusion than follows from the nonexistence of FPTAS for maximizing over $\mathrm{V}(k,n)$ in Theorem 7.2. □

Note that while Theorem 7.2 is proved for the bit model, Corollary 7.6 is proved in the oracle model. The reason is as in the commentary after the proof of Corollary 6.5 — note that θ_2 appears in every function in (9), and evaluating θ_2 involves a QR decomposition.

8. NP-hardness of QP on the Cartan manifold

The set of postive definite matrices $\mathbb{S}^n_{++}$ has two natural metrics — the standard Euclidean metric $\langle X, Y\rangle = \operatorname{tr}(XY)$ and the Riemannian metric $\mathsf{g}_A(X,Y) = \operatorname{tr}(A^{-1}XA^{-1}Y)$, defined for tangent vectors $X, Y \in \mathbb{S}^n$ and $A \in \mathbb{S}^n_{++}$. The former is rightly called the positive definite *cone* since a cone is, by definition, a subset of an Euclidean space and endowed with the standard Euclidean metric (which is in fact required to define a dual cone). When endowed with the latter metric, $(\mathbb{S}^n_{++}, \mathsf{g})$ is no longer a cone, but a natural model for $\mathcal{E}(\mathbb{R}^n)$, the set of all centered ellipsoids in $\mathbb{R}^n$, and g is the unique Riemannian metric that captures the geometry of ellipsoids.

This Riemannian manifold was first discussed by Élie Cartan at length in [3, pp. 364–372] and more briefly in [4]. There are many modern expositions, most notably Lang's [23] and [24, Chapter XII] but also Eberlein's [8, 9, 10] and Mostow's [28] and [29, Section 3]. All three authors credited [3, 4], directly or indirectly, for the most substantive features of this manifold, with Mostow explicitly attributing it to Cartan, and we follow suit in our article.

By way of the following lemma, we will deduce the section title from deciding copositivity of a matrix, one of the oldest NP-hard problems in optimization [31].

Lemma 8.1. *Let $A \in \mathbb{R}^{n\times n}$ and set*

$$\alpha_{ijkl} := \begin{cases} a_{ik} & i = j \text{ and } k = l, \\ 0 & \text{otherwise.} \end{cases}$$

for all $i, j, k, l \in \{1, \dots, n\}$. We define the quadratic form $f : \mathbb{S}^n \times \mathbb{S}^n \to \mathbb{R}$ by

$$f(X) = \sum_{i,j,k,l=1}^{n} \alpha_{ijkl} x_{ij} x_{kl}.$$

If $f(X) \geq 0$ for all $X \in \mathbb{S}^n_{++}$, then A is copositive.

Proof. This follows from

$$f(X) = \sum_{i,k=1}^{n} a_{ik} x_{ii} x_{kk} = x^{\mathsf{T}} A x$$

with $x := \operatorname{diag}(X)$. If $f(X) \geq 0$ for all $X \in \mathbb{S}^n_{++}$, then $f(X) \geq 0$ for all $X \in \mathbb{S}^n_{+}$. Note that as X runs over $\mathbb{S}^n_{+}$, its diagonal x runs over $\mathbb{R}^n_{+}$. □

Corollary 8.2 (Cartan quadratic programming is NP-hard)**.** *Unconstrained quadratic optimization over $\mathbb{S}^n_{++}$, whether regarded as the positive definite cone or the Cartan manifold, is NP-hard.*

Proof. If we can minimize f in Lemma 8.1 over $\mathbb{S}^n_{++}$, then we can decide if $f(X) \geq 0$ for all $X \in \mathbb{S}^n_{++}$, and thereby whether A is copositive. □

At this point it is worth highlighting the fact that the choice of metric does not play a role in NP-hardness results in optimization. The computational tractability or intractability of computing $\max_{x\in\mathcal{M}} f(x)$ depends only on $\mathcal{M}$ as a set. The choice of Riemannian metric does of course affect the convergence rate of the algorithm. For example, when minimizing a strictly convex $f : \mathbb{R}^n \to \mathbb{R}$, the usual gradient descent in the standard Euclidean metric $\langle v, w\rangle := v^{\mathsf{T}} w$ is linearly convergent, but gradient descent in the Riemannian metric $\langle v, w\rangle_x := v^{\mathsf{T}} \nabla^2 f(x)^{-1} w$ is equivalent to Newton's method, which is quadratically convergent. But no Riemannian metric we put on $\mathbb{R}^n$ will ever yield a polynomial-time algorithm for an NP-hard optimization problem unless P = NP.

We now prove an analogue of Proposition 6.1 required to extend Corollary 8.2 to $\operatorname{GL}(n)/\operatorname{O}(n)$, the quotient model of the Cartan manifold.

Lemma 8.3. *There exists a polynomial-time computable diffeomorphism* $\rho : \mathrm{GL}(n)/\,\mathrm{O}(n) \to \mathbb{S}^n_{++}$.

Proof. The required map and its inverse are given by

$$\rho : \mathrm{GL}(n)/\,\mathrm{O}(n) \to \mathbb{S}^n_{++}, \quad \rho(\llbracket X \rrbracket) = X^{\mathsf{T}} X,$$
$$\rho^{-1} : \mathbb{S}^n_{++} \to \mathrm{GL}(n)/\,\mathrm{O}(n), \quad \rho^{-1}(S) = \llbracket R_S \rrbracket,$$

where R_S is the unique upper-triangular matrix in the Cholesky decomposition $S = R_S^{\mathsf{T}} R_S$. It is routine to verify that ρ^{-1} is indeed the inverse of ρ; and it is computable in polynomial time since Cholesky decomposition is. That ρ is a diffeomorphism follows the same argument used in the proof of Proposition 6.1. □

For the next result, we will again have to resort to oracle complexity, as the commentary after the proof of Corollary 6.5 about QR decomposition also applies to Cholesky decomposition.

Corollary 8.4 (Cartan optimization is NP-hard)**.** *Unless* P = NP*, there is no oracle-*FPTAS *that is polynomial in n for maximizing real-valued functions over* $\mathrm{GL}(n)/\,\mathrm{O}(n)$.

Proof. By Lemma 8.3, ρ and ρ^{-1} are both polynomial-time computable. We may maximize the function $f : \mathbb{S}^n_{++} \to \mathbb{R}$ in Lemma 8.1 by querying the function $f \circ \rho : \mathrm{St}(k, n)/\,\mathrm{GL}(k) \to \mathbb{R}$ as an oracle and using ρ^{-1} to map a solution in $\mathrm{St}(k, n)/\,\mathrm{GL}(k)$ back to $\mathbb{S}^n_{++}$ in polynomial-time. By Corollary 8.2, it must be NP-hard to maximize over $\mathrm{GL}(n)/\,\mathrm{O}(n)$. □

9. LP is polynomial-time

The hardness results involving quadratic programming are the best possible in the sense that unconstrained linear programs are polynomial-time solvable for the Stiefel, Grassmann, and Cartan manifolds. They have closed-form solutions computable in polynomial time, easy exercises along the lines of those in [14], but recorded below for completeness.

Lemma 9.1 (Unconstrained LP over Stiefel, Grassmann, and Cartan manifolds)**.** (a) *For any* $A \in \mathbb{R}^{n \times k}$,

$$\max_{X \in \mathrm{V}(k,n)} \operatorname{tr}(A^{\mathsf{T}} X) = \sum_{i=1}^{k} \sigma_i$$

is attained at $X = UV^{\mathsf{T}}$ *where* $A = U\Sigma V^{\mathsf{T}}$ *is a condensed singular value decomposition with* $U, V \in \mathrm{V}(n, k)$, $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_k) \in \mathbb{R}^{k \times k}$, $\sigma_1 \ge \dots \ge \sigma_k \ge 0$.

(b) *For any* $A \in \mathbb{R}^{n \times n}$,

$$\max_{P \in \mathrm{Gr}(k,n)} \operatorname{tr}(A^{\mathsf{T}} P) = \sum_{i=1}^{k} \lambda_i$$

is attained at $P = Q_k Q_k^{\mathsf{T}}$ *where* $(A + A^{\mathsf{T}})/2 = Q \Lambda Q^{\mathsf{T}}$ *is an eigenvalue decomposition with* $Q \in \mathrm{O}(n)$, $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n \times n}$, $\lambda_1 \ge \dots \ge \lambda_n$.

(c) *For any* $A \in \mathbb{R}^{n \times n}$,

$$\sup_{X \in \mathbb{S}^n_{++}} \operatorname{tr}(A^{\mathsf{T}} X) = \begin{cases} 0 & \textit{if } -(A + A^{\mathsf{T}}) \in \mathbb{S}^n_{+}, \\ +\infty & \textit{otherwise.} \end{cases}$$

Proof. The problem in (a) transforms into

$$\max_{X \in \mathrm{V}(k,n)} \operatorname{tr}(A^{\mathsf{T}} X) = \max_{X \in \mathrm{V}(k,n)} \operatorname{tr}(\Sigma^{\mathsf{T}} X) = \max_{X \in \mathrm{V}(k,n)} \sum_{i=1}^{k} \sigma_i v_{ii}.$$

For any $X \in \mathrm{V}(k, n)$, the diagonal entries $x_{ii} \le 1$, so the maximum is bounded by $\sum_{i=1}^{k} \sigma_i$. On the other hand, $V = I_k$ achieves this bound.

The problem in (b) transforms into

$$\max_{P\in\operatorname{Gr}(k,n)} \operatorname{tr}(A^{\mathsf{T}}P) = \max_{P\in\operatorname{Gr}(k,n)} \operatorname{tr}\left(\left[\frac{A+A^{\mathsf{T}}}{2}\right]^{\mathsf{T}} P\right) = \max_{P\in\operatorname{Gr}(k,n)} \operatorname{tr}(\Lambda^{\mathsf{T}}P) = \max_{P\in\operatorname{Gr}(k,n)} \sum_{i=1}^{k} \lambda_i p_{ii}.$$

For any $P \in \operatorname{Gr}(k,n)$, the diagonal entries satisfy $0 \le p_{ii} \le 1$, $p_{11} + \cdots + p_{nn} = k$. Under these constraints, the function $\sum_{i=1}^{k} \lambda_i p_{ii}$ must be upper-bounded by $\sum_{i=1}^{k} \lambda_i$: If one of p_{ii} is not 1, we may always change it to 1 and decrease the value of some p_{ii}, $i > k$, and increase the function value. After at most k such changes, we arrive at the value $\sum_{i=1}^{k} \lambda_i$. On the other hand, $P = I_k$ achieves this bound.

For the problem in (c), if there is an $X \in \mathbb{S}^n_{++}$ with $\operatorname{tr}(A^{\mathsf{T}}X) > 0$, then $\operatorname{tr}(A^{\mathsf{T}}(\mu X)) = \mu \operatorname{tr}(A^{\mathsf{T}}X) \to +\infty$ as $\mu \to +\infty$. Otherwise, $\operatorname{tr}(A^{\mathsf{T}}X) \le 0$ for all $X \in \mathbb{S}^n_{++}$ and $X \to 0$ gives the least upper bound. □

10. Conclusion

In case one thinks that we might get around these intractability results by imposing convexity on the objective function. We remind the reader that for any compact manifold without boundary, which includes $\operatorname{Gr}(k,n)$, $\operatorname{V}(k,n)$, and $\operatorname{O}(n)$, the only continuous geodesically convex function is the constant function. Indeed this holds true for many manifolds that are noncompact as well [43].

Nevertheless, one positive way to view the hardness results in this article is that they show that current pursuits in manifold optimization, which are exclusively focused on local optimization, are most likely the right way to go. Short of having P = NP, manifold optimization cannot become a topic in global optimization like convex optimization.

Future work. Our constructions show that unconstrained cubic and quartic programming over $\operatorname{O}(n)$ and $\operatorname{V}(k,n)$ are NP-hard for all fixed $k \in \mathbb{N}$. For $k = 1$, quadratic programming over $\operatorname{V}(1,n)$ is well-known to be polynomial-time solvable. It is also common knowledge that the NP-hard quadratic assignment problem may be written as a linearly constrained quadratic optimization problem over $\operatorname{O}(n)$ [37]. These observations leave open the complexity of unconstrained quadratic programming over $\operatorname{O}(n)$ and $\operatorname{V}(k,n)$ for $k \ge 2$.

References

[1] P. Bürgisser, M. Clausen, and M. A. Shokrollahi. *Algebraic complexity theory*, volume 315 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1997.

[2] A. R. Calderbank, R. H. Hardin, E. M. Rains, P. W. Shor, and N. J. A. Sloane. A group-theoretic framework for the construction of packings in Grassmannian spaces. *J. Algebraic Combin.*, 9(2):129–140, 1999.

[3] E. Cartan. Sur certaines formes Riemanniennes remarquables des géométries à groupe fondamental simple. *Ann. Sci. École Norm. Sup. (3)*, 44:345–467, 1927.

[4] E. Cartan. Groupes simples clos et ouverts et géométrie riemannienne. *J. Math. Pures Appl. (9)*, 8(1):1–33, 1929.

[5] Y. Chikuse. *Statistics on special manifolds*, volume 174 of *Lecture Notes in Statistics*. Springer-Verlag, New York, NY, 2003.

[6] J. H. Conway, R. H. Hardin, and N. J. A. Sloane. Packing lines, planes, etc.: packings in Grassmannian spaces. *Experiment. Math.*, 5(2):139–159, 1996.

[7] E. de Klerk. The complexity of optimizing over a simplex, hypercube or sphere: a short survey. *CEJOR Cent. Eur. J. Oper. Res.*, 16(2):111–125, 2008.

[8] P. B. Eberlein. Structure of manifolds of nonpositive curvature. In *Global Differential Geometry and Global Analysis*, volume 1156 of *Lecture Notes in Math.*, pages 86–153. Springer, Berlin, 1984.

[9] P. B. Eberlein. Manifolds of nonpositive curvature. In *Global Differential Geometry*, volume 17 of *MAA Studies in Mathematics*, pages 223–258. Mathematical Association of America, Washington, DC, 1989.

[10] P. B. Eberlein. *Geometry of nonpositively curved manifolds*. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1996.

[11] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1999.

[12] M. Ehler and M. Gräf. Reproducing kernels for the irreducible components of polynomial spaces on unions of Grassmannians. *Constr. Approx.*, 49(1):29–58, 2019.

[13] G. P. Gehér and P. ˇ Semrl. Isometries of Grassmann spaces. *J. Funct. Anal.*, 270(4):1585–1601, 2016.

[14] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[15] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics: Study and Research Texts*. Springer-Verlag, Berlin, 1988.

[16] J. Harris. *Algebraic geometry*, volume 133 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, NY, 1995.

[17] S. Helgason. *Differential geometry, Lie groups, and symmetric spaces*, volume 34 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[18] A. Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *Amer. J. Math.*, 76:620–630, 1954.

[19] B. Jiang, Z. Li, and S. Zhang. On cones of nonnegative quartic forms. *Found. Comput. Math.*, 17(1):161–197, 2017.

[20] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972)*, The IBM Research Symposia Series, pages 85–103. Plenum, New York, NY, 1972.

[21] Z. Lai, L.-H. Lim, and K. Ye. Simpler Grassmannian optimization. *arXiv:2009.13502*, 2020.

[22] Z. Lai, L.-H. Lim, and K. Ye. Simple matrix expressions for the curvatures of Grassmannian. *arXiv:2406.11821*, 2024.

[23] S. Lang. Bruhat–Tits spaces. In *Math talks for undergraduates*, pages 79–100. Springer-Verlag, New York, NY, 1999.

[24] S. Lang. *Fundamentals of differential geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, NY, 1999.

[25] J. M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013.

[26] L.-H. Lim and K. Ye. Degree of the Grassmannian as an affine variety. *arXiv:2405.05128*, 2024.

[27] P. Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[28] G. D. Mostow. Some new decomposition theorems for semi-simple groups. *Mem. Amer. Math. Soc.*, 14:31–54, 1955.

[29] G. D. Mostow. *Strong rigidity of locally symmetric spaces*, volume 78 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1973.

[30] T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian J. Math.*, 17:533–540, 1965.

[31] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Math. Programming*, 39(2):117–129, 1987.

[32] A. S. Nemirovsky and D. B. Yudin. *Problem complexity and method efficiency in optimization*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons, Inc., New York, NY, 1983.

[33] Y. Nesterov. Random walk in a simplex and quadratic optimization over convex polytopes. Technical report, CORE, 2003.

[34] Y. Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, Cham, second edition, 2018.

[35] L. I. Nicolaescu. *Lectures on the geometry of manifolds*. World Scientific Publishing, Hackensack, NJ, third edition, 2021.

[36] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000.

[37] P. M. Pardalos, F. Rendl, and H. Wolkowicz. The quadratic assignment problem: a survey and recent developments. In *Quadratic assignment and related problems (New Brunswick, NJ, 1993)*, volume 16 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 1–42. American Mathematical Society, Providence, RI, 1994.

[38] A. Schönhage. Unitäre Transformationen grosser Matrizen. *Numer. Math.*, 20:409–417, 1972/73.

[39] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path of linear programs by linear extrapolation. II. volume 52, pages 527–553. 1991.

[40] E. Stiefel. Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten. *Comment. Math. Helv.*, 8(1):305–353, 1935.

[41] S. A. Vavasis. Quadratic programming is in NP. *Inform. Process. Lett.*, 36(2):73–77, 1990.

[42] S. A. Vavasis. *Nonlinear optimization*, volume 8 of *International Series of Monographs on Computer Science*. The Clarendon Press, Oxford University Press, New York, NY, 1991.

[43] S. T. Yau. Non-existence of continuous convex functions on certain Riemannian manifolds. *Math. Ann.*, 207:269–270, 1974.

[44] G. Zhao. Representing the space of linear programs as the Grassmann manifold. *Math. Program.*, 121(2, Ser. A):353–386, 2010.

Department of Mathematics, University of Texas, Austin, TX 78712
Email address: zehua.lai@austin.utexas.edu

Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago, Chicago, IL 60637
Email address: lekheng@uchicago.edu

KLMM, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
Email address: keyk@amss.ac.cn