



Elliptic polytopes and invariant norms of linear operators

Thomas Meistrick¹ · Valdimir Yu. Protasov²

Received: 20 February 2023 / Revised: 20 September 2023 / Accepted: 23 October 2023 /
Published online: 17 November 2023
© The Author(s) under exclusive licence to Istituto di Informatica e Telematica (IIT) 2023

Abstract

Elliptic polytopes are convex hulls of several concentric plane ellipses in \mathbb{R}^d . They arise in applications as natural generalizations of usual polytopes. In particular, they define invariant convex bodies of linear operators, optimal Lyapunov norms for linear dynamical systems, etc. To construct elliptic polytopes one needs to decide whether a given ellipse is contained in the convex hull of other ellipses. We analyse the computational complexity of this problem and show that for $d = 2, 3$, it admits an explicit solution. For larger d , two geometric methods for approximate solution are presented. Both use the convex optimization tools. The efficiency of the methods is demonstrated in two applications: the construction of extremal norms of linear operators and the computation of the joint spectral radius/Lyapunov exponent of a family of matrices.

Keywords Lyapunov function · Convex hull · Ellipse · Discrete time linear system · Joint spectral radius · Complexity

Mathematics Subject Classification 52A21 · 39A30 · 15A60 · 90C90

1 Introduction

Convex hulls of two-dimensional ellipses in \mathbb{R}^d can efficiently replace usual polytopes in some problems of numerical linear algebra and of the operator theory. They provide convenient tools in the evaluation of Lyapunov functions and of extremal norms of linear operators, in the study of stability of linear switching systems and in the

✉ Thomas Meistrick
tommsch@gmx.at

Valdimir Yu. Protasov
vladimir.protasov@univaq.it

¹ NuHAG, University of Vienna, Universitätsring 1, 1010 Vienna, Austria

² DISIM, University of L'Aquila, Piazza Santa Margherita 2, 67100 L'Aquila, Italy

computation of the joint spectral radius. However, the construction of such convex hulls is computationally hard, especially in high dimensions. It is reduced to the following question: to decide whether a given ellipse is contained in the convex hull of other given ellipses. We study the complexity of this problem, show that it admits an explicit solution in low dimensions, and for higher dimensions, we derive several methods of its approximate solution.

Note that a solution merely by approximating each ellipse with a polygon is inefficient and leads to enormous computations to achieve a good precision. That is why the problem requires other approaches based on various geometric ideas. The paper is concluded with numerical results and applications. We also establish the relation to the well-known concept of *balanced complex polytopes*.

Definition 1.1 An elliptic polytope in \mathbb{R}^d is a convex hull of several two-dimensional ellipses centred at the origin. Those ellipses which are not in the convex hull of the others are called vertices of the elliptic polytope.

An ellipse can be degenerate, in which case it is a segment centred at the origin. So, every (usual) polytope symmetric about the origin is also an elliptic polytope. We define an ellipse either by a pair of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ as the set of points $\mathbf{a} \cos t + \mathbf{b} \sin t$, $t \in \mathbb{R}$ and denote it as $E(\mathbf{a}, \mathbf{b})$, or by a complex vector $\mathbf{v} = \mathbf{a} + i\mathbf{b} \in \mathbb{C}^d$, and denote it as $E(\mathbf{v}) = E(\mathbf{a}, \mathbf{b})$, where $\mathbf{a} = \Re \mathbf{v}$, $\mathbf{b} = \Im \mathbf{v}$ are real and imaginary parts of \mathbf{v} , respectively.

Two complex vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{C}^d$ define the same ellipse if either $\mathbf{v}_2 = z\mathbf{v}_1$ or $\mathbf{v}_2 = z\bar{\mathbf{v}}_1$ for some $z \in \mathbb{C}$, $|z| = 1$.

The plan of the paper is the following. In the next section we discuss the main applications of elliptic polytopes in problems of linear algebra and linear dynamical systems. Then we formulate the main problem, describe the main results, discuss their novelty and the main methods of the study.

In Sect. 4 we give necessary definitions, notation, and formulate auxiliary facts. In Sect. 5 we reduce the main problems in the optimization problems and study their complexity. As we shall see, the problem is computationally hard.

Section 6 deals with low-dimensional case, where the problem can be explicitly solved. This fact is rather surprising, taking into account that the number of ellipses can be arbitrarily large. Then we present two methods of approximate solution for higher dimensions: the complex polytope method (Sect. 8) and the cutting angle method (Sects. 9 and 10). As we shall see, the latter can be significantly improved by implementing the high-dimensional projection approach [2] for solving quadratic programming problems.

Numerical results and illustrative examples are presented in Sect. 11 followed by examples from applications.

2 Motivation and applications

Construction of elliptic polytopes arise naturally in the study of spectral properties of matrices, of asymptotics of long matrix products, the stability of linear dynamical systems, and in related problems. Below we consider some of these applications.

Application 1. Norms in \mathbb{C}^d restricted to \mathbb{R}^d

It is well-known that every convex body in \mathbb{R}^d symmetric about the origin generates a norm in \mathbb{R}^d , called its Minkowski norm. In contrast, not every convex body in \mathbb{C}^d defines a norm in \mathbb{C}^d . Such bodies have a particular structure: If S is a unit sphere of a norm $\|\cdot\|$ in \mathbb{C}^d , then for every $\mathbf{v} \in S$, the curve $\{e^{-is}\mathbf{v} : s \in \mathbb{R}\}$ lies on S . Indeed, $\|e^{-is}\mathbf{v}\| = \|\mathbf{v}\| = 1$. Note that the real part of the point $\mathbf{v}(s)$ runs over the ellipse $E(\mathbf{v})$ as $s \in \mathbb{R}$. Thus, a unit ball of a norm in \mathbb{R}^d induced by an arbitrary complex norm is a convex hull of a (possibly infinite) set of ellipses. In particular, a piecewise linear approximation of a norm in the complex space is a *balanced complex polytope* with real and imaginary parts being elliptic polytopes. Therefore, elliptic polytopes are the real and imaginary part of a polyhedral approximation for unit balls of norms in \mathbb{C}^d .

Application 2. Lyapunov functions for linear dynamical systems

For a discrete time linear system of the form $\mathbf{x}(k + 1) = A\mathbf{x}(k)$, $k \in \mathbb{N}$, where A is a constant $d \times d$ matrix, an important issue is the construction of a Lyapunov function $f(\mathbf{x})$, for which $f(A\mathbf{x}) \leq \lambda f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$. If such a function exists for $\lambda = 1$, then the system is stable, if it exists for $\lambda < 1$, then it is asymptotically stable. A Lyapunov function provides a detailed information on the asymptotic behaviour of the trajectories $\mathbf{x}(k)$ as $k \rightarrow \infty$. A standard approach is to find a quadratic Lyapunov function $f(\mathbf{x}) = \sqrt{\mathbf{x}^T M \mathbf{x}}$, where M is a positive definite matrix. By the Lyapunov theorem such a matrix M exists whenever $\rho(A) < 1$, where ρ is the spectral radius (maximal modulus of eigenvalues). The quadratic Lyapunov function can be found either by solving a semidefinite programming problem, or by finding all eigenvectors of A . In high dimensions (larger than 25–100), however, both those methods become hard. In this case one should consider Lyapunov functions from other classes, for example, from the class of polyhedral functions. To construct a polyhedral Lyapunov function one needs to find a polytope P such that $AP \subset P$. Such a polytope can be constructed iteratively starting with an arbitrary polytope P_0 and running the process $P_{k+1} = \text{co}\{AP_k, P_k\}$, where co denotes the convex hull. When $P_{n+1} = P_n$ the algorithm halts and we set $P = P_n$. In general, this algorithm never stops except when started with the leading eigenvector(s) \mathbf{v} related to the problem. If \mathbf{v} is not real then P_0 should be chosen as $E(\mathbf{v})$ and the final polytope P will be elliptic.

Application 3. Computation of the joint spectral radius

This is one of the most important applications of elliptic polytopes. The joint spectral radius of matrices is the maximal rate of asymptotic growth of norms of their long products. For an arbitrary family $\mathcal{A} = \{A_1, \dots, A_m\}$ of $d \times d$ matrices, the joint spectral radius (JSR) is the limit

$$\rho(\mathcal{A}) = \lim_{k \rightarrow \infty} \max_{A(j) \in \mathcal{A}} \|A(k) \dots A(1)\|^{1/k}. \tag{1}$$

Originated with J. K. Rota and G. Strang in 1960 the joint spectral radius found numerous applications, see [3, 9, 19, 21, 22, 29] for surveys. The computation of the joint spectral radius, even its approximation, is a hard problem.

The *Invariant polytope algorithm* introduced in [9] makes it possible to find a precise value of $\rho(\mathcal{A})$ for a vast majority of matrix families. Its idea traces back to [11, 27]. Recent works [10, 23, 28] develop updated versions of that algorithm which efficiently perform computations in dimensions up to 25 for arbitrary matrices and up to several thousands for non-negative matrices. The main idea of the Invariant polytope algorithm is the following:

First, we find a candidate for the spectrum maximizing product Π of matrices from \mathcal{A} , for which the largest eigenvalue in modulus (the *leading eigenvalue*) λ is maximal, i.e. $|\lambda| = \rho(\Pi)^{1/|\Pi|}$.

We make the assumption that the leading eigenvalue of Π is unique and simple. Then we construct an *extremal norm* $\|\cdot\|$ in \mathbb{R}^d such that $\|A\mathbf{x}\| \leq |\lambda| \|\mathbf{x}\|$ for all $A \in \mathcal{A}$, $\mathbf{x} \in \mathbb{R}^d$. Once such a norm is found, we have proved that $\rho(\mathcal{A}) = |\lambda|$.

If $\lambda \in \mathbb{R}$, then the extremal norm is constructed iteratively, starting with the leading eigenvector \mathbf{v} of Π (i.e. the eigenvector corresponding to the leading eigenvalue), considering its m images $\lambda^{-1}A\mathbf{v}$, $A \in \mathcal{A}$, then constructing their m^2 images, etc.. To avoid the exponential growth of the number of points we remove all redundant points (those in the convex hull of others) in each iteration. If this process halts after several iterations (no new points appear), then the convex hull of the collected points forms an *invariant polytope* P , for which $AP \subset \lambda P$ for all $A \in \mathcal{A}$. The Minkowski norm of this polytope is extremal.

If the leading eigenvalue of Π is non-real, then P can be found as a balanced complex polytope (Definition 4.1) by the same iterative procedure starting with complex leading eigenvector \mathbf{v} . This approach was elaborated in [11–13] and shown to be efficient for the JSR computation. There were, however, some disadvantages. First of all, this method was mostly applied in low dimensions. Second, for matrix families with complex leading eigenvalue

this algorithm suffers, since the uniqueness of the leading eigenvalue assumption is violated. Indeed for the latter case, if λ is a leading eigenvalue, then $\bar{\lambda}$ is also a leading entry.

We modify this method by using elliptic polytopes instead of balanced complex polytopes.

Application 4. Stability of linear switching systems

Extremal norm $\|\cdot\|$ constructed by means of elliptic polytopes for computing the joint spectral radius play not solely an auxiliary role. They are of an independent interest as Lyapunov functions for the discrete time *linear switching system* $\mathbf{x}(k+1) = A(k)\mathbf{x}(k)$, $A(k) \in \mathcal{A}$, $k \geq 0$, see [1, 6, 15, 20, 26, 31, 32]. For this system, $\rho(\mathcal{A})$ has the meaning of the Lyapunov exponent, and the extremal norm $\|\cdot\|$ is a Lyapunov function. Thus, the Lyapunov function of a discrete time linear switching system is constructed as the Minkowski functional of an elliptic polytope.

3 Problem statement and the description of main results

The following problem, which will be referred to as Problem **EE** (*ellipse in ellipses*) is crucial in constructing and studying elliptic polytopes.

Problem EE For given ellipses E_0, \dots, E_N in \mathbb{R}^d , decide whether $E_0 \subset \text{co}\{E_1, \dots, E_N\}$.

An efficient solution to Problem **EE** makes it possible to “clean” every set of ellipses removing redundant ones and leaving only the vertices of an elliptic polytope containing all others. All the aforementioned applications in Sect. 1 are based on the use of Problem **EE**.

Concerning Application 1, an arbitrary norm in \mathbb{C}^d can be approximated by a polyhedral norm $\|\mathbf{x}\| = \max_j |(v_j, \mathbf{x})|$. Consider the restriction of this norm to \mathbb{R}^d . Solving Problem **EE** one removes redundant vectors v_k . A term $|(v_k, \mathbf{x})|$ is redundant if and only if the ellipse $E(v_k)$ is contained in the convex hull of the others $E(v_j)$, $j \neq k$.

In the other applications, solving Problem **EE** also plays a major role. In the iterative construction of the Lyapunov functions and in the Invariant polytope algorithms, the removal of redundant ellipses in each iteration prevents the exponential growth of the number of ellipses and actually makes those algorithms applicable. Moreover, reducing the number of ellipses makes the Lyapunov function simpler and more convenient for applications.

Our second topic is the algorithmic implementation of the solution of Problem **EE**. In particular, we aim to modify the algorithm of the joint spectral radius (JSR) computation (Application 3) by using elliptic polytopes instead of complex polytopes. The same construction will be applied for finding invariant Lyapunov functions for switching systems (Application 4).

3.1 Possible approaches

An analogue to Problem **EE** for usual polytopes is solved by the standard linear programming technique. For elliptic polytopes, we are not aware of any known method. To the best of our knowledge, the only problem considered in the literature, which is related to Problem **EE**, is the construction of a balanced complex polytope. This technique was developed in [9, 11–14] for finding extremal Lyapunov functions in \mathbb{C}^d and for computing the joint spectral radius. It is based on the following fact: An ellipse $E(v_0)$ is contained in the convex hull $\text{co}\{E(v_1), \dots, E(v_N)\}$ if there exist complex numbers z_k such that $v_0 = \sum_{k=1}^m z_k v_k$ and $\sum_{k=1}^m |z_k| \leq 1$. This condition, however, is only sufficient but not necessary. Moreover, it turns out that for solving Problem **EE**, this method gives a rather rough approximate solution. We are going to show that its approximation factor is $1/2$ and this estimate is tight (Theorems 8.2 and 8.3 in Sect. 7). Moreover, it works only if we add the complex conjugate vectors \bar{v}_k to the set of vectors v_k , otherwise the approximation factor is zero. I.e. we will not obtain even an approximate solution. This aspect has been missed in the recent literature on the joint spectral radius computation.

Natural questions arise—How to get a precise solution of Problem EE and what is the complexity of this problem? What could be done to obtain approximate solutions with better approximation factors? Having answered those questions one can speed up the Invariant polytope algorithm for the joint spectral radius computation, construct extremal Lyapunov functions for discrete time systems that would be easier to define and to compute than those presented in the literature, and address other applications.

3.2 Summary of the main results

We begin with reducing Problem EE to an optimization problem (Sect. 5) The problem is highly non-convex and, therefore, can be hard. Indeed, we show that it is not simpler than the problem of maximizing a quadratic form of rank 2 over a centrally symmetric polyhedron. We conjecture that the latter problem is NP-hard. An argument for that is established in Theorem 5.4, a positive semidefinite quadratic form of rank 2 in \mathbb{R}^k under $O(k)$ linear constraints may have 2^k points of local maxima.

In low dimensions Problem EE admits a precise solution. This is shown in Sect. 6. Moreover, if the dimension is fixed, then the problem has a polynomial (in the number of ellipses) solution, although hardly realizable for $d \geq 4$. For higher dimensions we can deal with approximate solutions only (Sect. 7).

The first method we present for the approximate solution of Problem EE is the *complex polytope method*. The theory of balanced complex polytopes originated with Guglielmi, Zennaro, and Wirth [11–13]. This method reduces Problem EE to a conic programming problem. First we observe one aspect missed in the literature: This method does not work, unless we add complex conjugate vectors to all given vectors (Proposition 8.1). After this modification, the method becomes applicable and gives an approximate solution to Problem EE with an approximation factor $1/2$. This is shown in Theorem 8.2. This factor is sharp and in general cannot be improved as shown in Theorem 8.3. Nevertheless, the empirical estimate obtained for random elliptic polytopes gives the expected value of the approximation factor around $1/\sqrt{2}$, which is, however, also quite rough. The main advantage of the complex polytope method is that it is simple and computationally cheap. The corresponding numerical results are presented in Sect. 11.

4. Our second method gives approximate solutions with a factor arbitrarily close to one (the factor 1 corresponds to the precise solution). This is a corner cutting algorithm derived in Sect. 9. By solving k conic programming problems with N constraints, where N is the number of ellipses, we get an approximate solution with an approximate factor of $1 - \pi^2/2(k+1)^2$. Already for $k = 3$, we obtain the factor at least $1/\sqrt{2} \simeq 0.707$, which is better than the one by the polytope method. For $k = 5$, the factor is approximately 0.923. These are the “worst case estimates” and in practice the corner cutting algorithm reaches a much higher accuracy already for small k .

In Sect. 10 we realize a modification of the corner cutting algorithm to a linear programming problem. This is done by applying the idea of Ben-Tal and Nemirovski of approximating quadrics by projections of higher dimensional polyhedra. The modified method reaches a very high accuracy.

After numerical results presented in Sect. 11 we illustrate some applications. First of all, the elliptic polytopes allow us to efficiently construct Lyapunov functions for linear dynamical systems even in high dimensions, for which a traditional way of finding a quadratic Lyapunov function by semidefinite programming is hardly reachable. Second, for the computation of the joint spectral radius of matrices, our results speed up the Invariant polytope algorithm in case of non-real leading eigenvalue and reduce a lot the number of ellipses defining the extremal Lyapunov function of the system. Finally, in the linear switching systems, the elliptic polytopes provide the invariant (Barabanov) norms.

4 Preliminary facts and notation

Throughout the paper we denote vectors by bold letters and numbers by standard letters. Thus $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$. As usual, for two complex vectors $\mathbf{v}, \mathbf{u} \in \mathbb{C}^d$, their scalar product is $(\mathbf{v}, \mathbf{u}) = \sum_{k=1}^d v_k \bar{u}_k$. For two real vectors \mathbf{a}, \mathbf{b} , we consider the ellipse

$$E = E(\mathbf{a}, \mathbf{b}) = \{ \mathbf{a} \cos t + \mathbf{b} \sin t : t \in \mathbb{R} \}. \tag{2}$$

This is an ellipse with conjugate radii (the halves of conjugate diameters) \mathbf{a}, \mathbf{b} . For a complex vector $\mathbf{v} = \mathbf{a} + i\mathbf{b}$ with real \mathbf{a} and \mathbf{b} , we write $E(\mathbf{v}) = E(\mathbf{a}, \mathbf{b})$. For every $s \in \mathbb{R}$, the real and complex parts of the vector $e^{-is}\mathbf{v} = \mathbf{a}_s - i\mathbf{b}_s$ are conjugate directions of the same ellipse and, therefore, $E(\mathbf{a}_s, \mathbf{b}_s) = E(\mathbf{a}, \mathbf{b})$ for all $s \in \mathbb{R}$. Indeed,

$$e^{-is}\mathbf{v} = (\cos s - i \sin s)(\mathbf{a} + i\mathbf{b}) = (\mathbf{a} \cos s + \mathbf{b} \sin s) + i(-\mathbf{a} \sin s + \mathbf{b} \cos s),$$

hence, $\mathbf{a}_s = \mathbf{a} \cos s + \mathbf{b} \sin s$ and $\mathbf{b}_s = \mathbf{a} \sin s - \mathbf{b} \cos s$. Therefore, $\mathbf{a}_s \cos t + \mathbf{b}_s \sin t = \mathbf{a} \cos(t + s) + \mathbf{b} \sin(t + s)$. The pair $(\mathbf{a}_s, \mathbf{b}_s)$ is the image of (\mathbf{a}, \mathbf{b}) after the elliptic rotation by the angle s along the ellipse $E = E(\mathbf{a}, \mathbf{b})$. Consequently, the vectors $\mathbf{a}_s, \mathbf{b}_s$ are also conjugate directions of the ellipse E .

Elliptic polytopes are real parts of the so-called balanced complex polytopes defined as follows:

Definition 4.1 A balanced convex hull of a set $K \subset \mathbb{C}^d$ is

$$\text{cob}(K) = \left\{ \sum_{k=1}^n z_k \mathbf{v}_k : z_k \in \mathbb{C}, \mathbf{v}_k \in K, \sum_{k=1}^n |z_k| \leq 1, n \in \mathbb{N} \right\}. \tag{3}$$

A balanced convex set is a subset of \mathbb{C}^d that coincides with its balanced convex hull, a balanced complex polytope is the balanced convex hull of a finite set of points.

If G is a balanced complex polytope, then ReG is a convex hull of ellipses. Indeed, if $G = \text{cob} \{ \mathbf{v}_1, \dots, \mathbf{v}_N \}$ and $\mathbf{a}_k = Re\mathbf{v}_k, \mathbf{b}_k = \Im \mathbf{v}_k, k = 1, \dots, N$, then for arbitrary complex numbers $z_k = r_k e^{-it_k}$, where $r_k = |z_k|, k = 1, \dots, N$, we have

$Re z_k \mathbf{v}_k = r_k (\mathbf{a}_k \cos t_k + \mathbf{b}_k \sin t_k)$, and hence, the set ReG consists precisely of the points $\sum_k r_k \mathbf{u}_k$ with $\mathbf{u}_k \in E(\mathbf{v}_k)$ and $\sum_k r_k \leq 1$. This is $\text{co} \{E(\mathbf{v}_1), \dots, E(\mathbf{v}_N)\}$.

Note that the balanced polytopes G and $\tilde{G} = \{\tilde{\mathbf{v}} : \mathbf{v} \in G\}$ have the same real parts and, hence, generate the same elliptic polytope P . Therefore, P does not change if we replace G by $\text{cob} \{G, \tilde{G}\}$. In what follows, if the converse is not stated, we assume that the balanced complex polytope is symmetric with respect to its conjugate, i.e. $G = \tilde{G}$. Clearly, this holds when the set of vertices is self-conjugated.

Given $P \subset \mathbb{R}^s$, compact, convex, with non-empty interior and such that $tP \subset P$ for all $|t| \leq 1$; We denote by $\|\cdot\|_P$ the vector norm whose unit ball is P .

Remark 4.2 The imaginary part of a balanced complex polytope is the same elliptic polytope P . Indeed,

$$\begin{aligned} \Im z_k \mathbf{v}_k &= r_k (-\mathbf{a}_k \sin t_k + \mathbf{b}_k \cos t_k) = r_k \left(\mathbf{a}_k \cos \left(t_k + \frac{\pi}{2} \right) + \mathbf{b}_k \sin \left(t_k + \frac{\pi}{2} \right) \right) \\ &= Re \tilde{z}_k \mathbf{v}_k, \end{aligned}$$

for some $\tilde{z}_k \in \mathbb{C}^s$. We see that the set $\Im G$ consists of the points $\sum_k r_k \mathbf{v}_k$ with $\mathbf{v}_k \in E(\mathbf{v}_k) =: E_k$ and $\sum_k r_k \leq 1$, and thus, $\Im G = P = ReG$. Of course, the same is true for an arbitrary balanced convex set: its real and imaginary parts coincide.

5 Equivalent optimization problems and their complexity

To analyse the complexity and possible solutions of Problem [EE](#) we reformulate it as an optimization problem.

5.1 Reformulation of Problem [EE](#)

Let $P = \text{co} \{E_1, \dots, E_N\}$ be an elliptic polytope. An ellipsoid E_0 is not contained in P if and only if P possesses a hyperplane of support that intersects E_0 at two points. For the outward normal vector $\mathbf{x} \in \mathbb{R}^d$ of that hyperplane, we have

$$\sup_{\mathbf{w}_0 \in E(\mathbf{a}_0, \mathbf{b}_0)} (\mathbf{x}, \mathbf{w}_0) > \sup_{\mathbf{w} \in P} (\mathbf{x}, \mathbf{w}).$$

Note that

$$\sup_{\mathbf{w} \in E(\mathbf{a}, \mathbf{b})} (\mathbf{x}, \mathbf{w}) = \sup_{t \in \mathbb{R}} (\mathbf{x}, \mathbf{a}) \cos t + (\mathbf{x}, \mathbf{b}) \sin t = \sqrt{(\mathbf{x}, \mathbf{a})^2 + (\mathbf{x}, \mathbf{b})^2}.$$

Therefore,

$$\begin{aligned} \sup_{\mathbf{w}_0 \in E(\mathbf{a}_0, \mathbf{b}_0)} (\mathbf{x}, \mathbf{w}_0) &= \sqrt{(\mathbf{x}, \mathbf{a}_0)^2 + (\mathbf{x}, \mathbf{b}_0)^2} \\ \sup_{\mathbf{w} \in P} (\mathbf{x}, \mathbf{w}) &= \max_{k=1, \dots, N} \sqrt{(\mathbf{x}, \mathbf{a}_k)^2 + (\mathbf{x}, \mathbf{b}_k)^2}. \end{aligned}$$

Thus, the assertion $E_0 \not\subset P$ is equivalent to the existence of a solution $\mathbf{x} \in \mathbb{R}^d$ for the system of inequalities

$$(\mathbf{x}, \mathbf{a}_0)^2 + (\mathbf{x}, \mathbf{b}_0)^2 > (\mathbf{x}, \mathbf{a}_k)^2 + (\mathbf{x}, \mathbf{b}_k)^2, \quad k = 1, \dots, N. \tag{4}$$

Normalizing the vector \mathbf{x} , it can be assumed that $(\mathbf{x}, \mathbf{a}_0)^2 + (\mathbf{x}, \mathbf{b}_0)^2 = 1 - \varepsilon$ where $\varepsilon > 0$ is a small number, in which case the system (4) is equivalent to the system $(\mathbf{x}, \mathbf{a}_k)^2 + (\mathbf{x}, \mathbf{b}_k)^2 \leq 1, k = 1, \dots, N$. Thus, we have proved:

Theorem 5.1 *Problem EE is equivalent to the following optimization problem:*

$$\begin{cases} (\mathbf{x}, \mathbf{a}_0)^2 + (\mathbf{x}, \mathbf{b}_0)^2 \rightarrow \max \\ (\mathbf{x}, \mathbf{a}_k)^2 + (\mathbf{x}, \mathbf{b}_k)^2 \leq 1, \quad k = 1, \dots, N, \end{cases} \tag{5}$$

with d variables $(x_1, \dots, x_d)^T = \mathbf{x}$ and given vectors $\mathbf{a}_k, \mathbf{b}_k \in \mathbb{R}^d$.

Therefore, we need to maximize a positive semidefinite quadratic form of rank two on the intersection of cylinders. Problem EE is hard since it involves a maximization; Even though all functions involved are convex, and the feasible set is convex.

5.2 The complexity of Problem EE

Maximization of a convex function over a convex set is usually non-trivial. Problem EE, and its reformulation (5), do not seem to be an exception. Moreover, the feasible domain is defined by N quadratic inequalities in \mathbb{R}^d and does not look simple either. Geometrically this is an intersection of N elliptic cylinders in \mathbb{R}^d with two-dimensional bases. The following result sheds some light on the complexity of this problem and, hence, on the complexity of Problem EE.

Theorem 5.2 *Maximization of a positive semidefinite quadratic form of rank two over a centrally symmetric polyhedron defined by $2N$ linear inequalities in \mathbb{R}^d can be reduced to Problem EE.*

Proof An origin-symmetric polyhedron is defined by N inequalities of the form $(\mathbf{x}, \mathbf{h}_k)^2 \leq 1$. Choosing arbitrary numbers $t_1, \dots, t_N \in (0, \frac{\pi}{2})$, we set $\mathbf{a}_k = \mathbf{h}_k \cos t_k, \mathbf{b}_k = \mathbf{h}_k \sin t_k$. Then the polytope is defined by the system of constraints of the reformulation (5). Finally, every quadratic form of rank two can be written as $(\mathbf{x}, \mathbf{a}_0)^2 + (\mathbf{x}, \mathbf{b}_0)^2$ for suitable $\mathbf{a}_0, \mathbf{b}_0$, which completes the proof. \square

Conjecture 5.3 *Maximizing a positive semidefinite quadratic form of rank two over a centrally symmetric polyhedron is NP-hard.*

Let us recall that the problem of maximizing a positive semidefinite quadratic form over a polyhedron is NP-hard even if that polyhedron is a unit cube, since it is not simpler than the Max-Cut problem [7, 16]. Moreover, even its approximate solution is NP-hard [17]. However, the rank two assumption may significantly simplify it. For example, the complexity of this problem on the unit cube becomes not only polynomial,

but linear with respect to Nd . It is reduced to finding the diameter of a flat zonotope, see [5] for more result on this and related problems. Nevertheless, we believe in the high complexity of this problem. One argument for that is a large number of local extrema. The following theorem states that if we drop the assumption of the symmetry of the polyhedron, then the number of local maxima with different values of the function can be exponential.

Theorem 5.4 *For each $N \geq 2$ there exists a polyhedron in \mathbb{R}^N with less than $2N$ facets and a positive semidefinite quadratic form of rank two on that polyhedron which has at least 2^{N-2} points of local maxima with different values of the function.*

The proof is in the Appendix.

On the other hand, as we shall see in the next section, in low dimensions Problem EE admits efficient solutions.

6 Problem EE in low dimensions

In dimensions $d = 2, 3$, Problem EE can be efficiently solved. The solution in the two-dimensional case is simple, the three-dimensional case is computationally harder.

6.1 The dimension $d = 2$

In the two-dimensional plane the solvability of the system (4) is explicitly decidable, which solves Problem EE.

Proposition 6.1 *In the case $d = 2$, Problem EE admits an explicit solution for arbitrary ellipses E_0, \dots, E_N . The complexity of the solution is linear in N .*

The proof is constructive and gives the method for the solution.

Proof Denote $\mathbf{x} = (x, y)^T$ and rewrite the inequalities (4) in coordinates. After simplifications we get $A_k y^2 + 2B_k xy + C_k x^2 > 0, k = 1, \dots, N$, where A_k, B_k, C_k are known coefficients. The set of solutions to the k th inequality is $\frac{y}{x} \in I_k$, where I_k is either the interval with ends at the roots of the quadratic equation $A_k t^2 + 2B_k t + C_k = 0$, if $A_k < 0$ (if there are no real roots, then $I_k = \emptyset$); or the union of two open rays with the same roots, if $A_k > 0$ (if there are no real roots, then $I_k = \mathbb{R}$); or one ray if $A_k = 0, B_k \neq 0$; the other cases are trivial. Then the solution of the system (4) consists of points $\mathbf{x} = (x, y)^T$ such that the ratio $\frac{y}{x}$ belongs to the intersection $\bigcap_{k=1, \dots, N} I_k$. Hence, $E_0 \subset \text{co} \{E_1, \dots, E_N\}$ if and only if this intersection is empty, i.e. $\bigcap_{k=1, \dots, N} I_k = \emptyset$. \square

6.2 The three dimensional case

In three dimensions the solvability of the system (4) is also explicitly decidable, but much harder than in dimension 2.

Proposition 6.2 For $d = 3$ and arbitrary ellipses E_0, \dots, E_N , Problem **EE** is reduced to solving of $O(N^2)$ bivariate quadratic systems of equations.

Proof Denote $\mathbf{x} = (x, y, z)^T$ and rewrite the inequalities (4) in coordinates. This is a system of homogeneous inequalities of degree 2. After the division by z^2 , we get a system of quadratic inequalities $f_i(x, y) < 0, i = 1, \dots, N$. It is compatible precisely when so is the system $f_i(x, y) - \varepsilon \leq 0, i = 1, \dots, N$, for some small $\varepsilon > 0$. Denote by \mathcal{D} the set of its solutions and assume it is non-empty. This is a closed subset of \mathbb{R}^2 bounded by arcs of the quadrics $\Gamma_i = \{(x, y)^T \in \mathbb{R}^2 : f_i(x, y) - \varepsilon = 0\}$. The closest to the origin point of \mathcal{D} belongs to one of the three sets: 1) the origin itself; 2) points of pairwise intersections $\Gamma_i \cap \Gamma_j, i \neq j$; 3) closest to the origin points of $\Gamma_i, i = 1, \dots, N$. If some of those quadrics coincide or are circles centred at the origin, then we reduce the number of quadrics by the standard argument. Otherwise, the set 2 contains at most $4 \cdot \frac{N(N-1)}{2} = 2N(N - 1)$ points; the set 3 contains at most $4N$ points. Hence, if \mathcal{D} is non-empty, then it contains one of the points of the sets 1, 2, 3. Thus, to decide if \mathcal{D} is empty or not, one needs to take each of those $2N(N - 1) + 4N + 1 = 2N^2 + 2N + 1$ points and check whether it belongs to \mathcal{D} , i.e. satisfies all the inequalities $f_i(x, y) - \varepsilon \leq 0, i = 1, \dots, N$. If the answer is affirmative for at least one point, then system (4) is compatible and $E_0 \not\subset P$, otherwise $E_0 \subset P$.

Evaluating each of those $O(N^2)$ points, except for the first one, is done by solving a system of two quadratic inequalities. □

6.3 Problem EE in a fixed dimension

Similarly to Proposition 6.2, one can show that Problem **EE** in \mathbb{R}^d is reduced to $O(N^{d-1})$ systems of d quadratic equations with d variables. The complexity of this problem is formally polynomial in N , with the degree depending on d . However the method used in the case $d = 3$ (the exhaustion of points of intersections and of points minimizing the distance to the origin) becomes non-practical for higher dimensions.

7 Approximate solutions

Apart from the low-dimensional cases, most likely, no efficient algorithms exist to obtain an explicit solution of Problem **EE**. That is why we are interested in approximate solutions with a given relative error (approximation factor) according to the following definition:

Definition 7.1 A method solves Problem **EE** approximately with a factor $q \in [0, 1]$ if it decides between two cases: either $E_0 \not\subset P$ or $qE_0 \subset P$.

So, the extreme case $q = 1$ corresponds to a precise solution, the other extreme case $q = 0$ means that the method does not give any approximate solution. We consider two methods. The first one is based on the construction of a balanced complex polytope. Such polytopes were deeply analysed in [11, 13, 14]. At the first sight, the results of those works give a straightforward solution to Problem **EE**. However, this is not the case. We are going to show that the balanced complex polytope method provides only

an approximate solution with the factor $q = 1/2$ and this value cannot be improved. Moreover, this approximation is attained only after a slight modification of this method, otherwise the approximation factor may drop to zero. Then we introduce the second method which provides a better approximation (with the factor q arbitrarily close to 1, i.e. to the precise solution).

8 The complex polytope method

Given an elliptic polytope $P = \text{co}\{E_1, \dots, E_N\}$ and an ellipse E_0 ; We need to decide whether $E_0 \subset P$. For each ellipse E_k , we choose arbitrary conjugate radii $\mathbf{a}_k, \mathbf{b}_k$, thus $E_k = E_k(\mathbf{a}_k, \mathbf{b}_k)$, $k = 1, \dots, N$. Define $\mathbf{v}_k = \mathbf{a}_k + i\mathbf{b}_k$ and consider the balanced complex polytope

$$G = \text{cob}\{\mathbf{v}_k : k = 1, \dots, N\}. \quad (6)$$

To get an approximate solution of Problem [EE](#) we consider the following auxiliary problem:

Problem [EE*](#) For points $\mathbf{v}_0, \dots, \mathbf{v}_N \in \mathbb{C}^d$, decide whether or not $\mathbf{v}_0 \in \text{cob}\{\mathbf{v}_1, \dots, \mathbf{v}_N\}$.

What is the relation between Problem [EE](#) and [EE*](#)? Clearly, if $\mathbf{v}_0 \in G$, then $E_0 \subset P$. Indeed, if $\mathbf{v}_0 \in G$, then $e^{it}\mathbf{v}_0 \in G$ for all $t \in \mathbb{R}$, hence $E_0 = \text{Re}\{e^{it}\mathbf{v}_0, t \in \mathbb{R}\} \subset \text{Re}G$. However, the converse is, in general, not true and Problem [EE*](#) is not equivalent to Problem [EE](#). Moreover, Problem [EE*](#) does not even provide an approximate solution to Problem [EE](#) with a positive factor. This means that the assertion $E_0 \subset P$ does not imply the existence of a positive q such that $q\mathbf{v}_0 \in G$.

Proposition 8.1 *Problem [EE*](#) gives an approximate solution to Problem [EE](#) with the factor $q = 0$.*

Proof Let $\mathbf{a}_0 = (1, 0)^T$ and $\mathbf{b}_0 = (0, 1)^T$, and $\mathbf{a}_1 = \mathbf{b}_0, \mathbf{b}_1 = \mathbf{a}_0$. Clearly, E_0 and E_1 both coincide with the unit disc, hence P is also a unit disc and $E_0 \subset P$. On the other hand, no positive number q exists such that $q\mathbf{v}_0 \in G$. Indeed, $G = \{z\mathbf{v}_1 : |z| \leq 1\}$. Denote $z = t + iu$. If $q\mathbf{v}_0 = z\mathbf{v}_1$, then $q\mathbf{v}_0 = t\mathbf{a}_1 - u\mathbf{b}_1 + i(t\mathbf{b}_1 + u\mathbf{a}_1) = t\mathbf{b}_0 - u\mathbf{a}_0 + i(t\mathbf{a}_0 + u\mathbf{b}_0)$. Hence, $q\mathbf{a}_0 = t\mathbf{b}_0 - u\mathbf{a}_0$ and $q\mathbf{b}_0 = t\mathbf{a}_0 + u\mathbf{b}_0$, which is coordinate wise $(q, 0) = (-u, t)$ and $(0, q) = (t, u)$. Therefore, $q = t = u = 0$. \square

Thus, Problem [EE*](#) does not give an approximate solution to Problem [EE](#). Nevertheless, under an extra assumption that G is self-conjugate, it does provide an approximate solution with the factor $1/2$. This factor is tight and cannot be improved. This follows from Theorems [8.2](#) and [8.3](#) proved below. Before formulating them, we briefly discuss the practical issue.

To solve Problem [EE*](#) we consider a self-conjugate balanced complex polytope $G = \text{cob}\{\mathbf{v}_k, \bar{\mathbf{v}}_k : k = 1, \dots, N\}$. As we noted in Remark [4.2](#), it has the same real

part P as the balanced polytope $G = \text{cob} \{v_k : k = 1, \dots, N\}$. Problem **EE*** is solved for G by the following optimization problem:

$$\left\{ \begin{array}{l} t_0 \rightarrow \max, \quad \text{subject to:} \\ \sqrt{t_j^2 + u_j^2} \leq r_j, \quad j = 1, \dots, 2N \\ \sum_{j=1}^{2N} r_j \leq 1 \\ t_0 \mathbf{a}_0 = \sum_{k=1}^N (t_k \mathbf{a}_k - u_k \mathbf{b}_k) + (t_{k+N} \mathbf{a}_k + u_{k+N} \mathbf{b}_k) \\ t_0 \mathbf{b}_0 = \sum_{i=1}^{\ell} (u_k \mathbf{a}_k + t_k \mathbf{b}_k) + (u_{k+N} \mathbf{a}_k - t_{k+N} \mathbf{b}_k) \end{array} \right. \tag{7}$$

This problem finds the biggest t_0 such that $t_0 v_0$ is a balanced complex combination of the points $v_1, \dots, v_N, \bar{v}_1, \dots, \bar{v}_N$. The coefficients of this combination are $z_k = t_k + i u_k, k = 1, \dots, 2N$, the points v_k, \bar{v}_k correspond to the coefficients z_k, z_{k+N} respectively. This is a convex conic programming problem with variables t_0, t_k, u_k , where $k = 1, \dots, 2N$. It is solved by the interior point method on Lorentz cones. If $t_0 \geq 1$, then $v_0 \in G$ and vice versa.

In Sect. 11 we present the numerical results showing that the problem is efficiently solved in relatively low dimension 2 to 20 and for the number of ellipses up to 2000.

Now we are going to see that if $v_0 \notin G$, then $E_0 \not\subset \frac{1}{2}P$. Dividing by two, we obtain an approximate solution to Problem **EE** with the factor at least $1/2$: if $\frac{1}{2}v_0 \notin G$, then $E_0 \not\subset P$, otherwise, if $\frac{1}{2}v_0 \in G$, then $\frac{1}{2}E_0 \subset P$.

Theorem 8.2 *If the set of vertices G is self-conjugated, then a precise solution of Problem **EE*** gives an approximate solution to Problem **EE** with the factor $q \geq \frac{1}{2}$.*

Proof It suffices to show that if $v_0 \notin G$, then $E_0 \not\subset \frac{1}{2}P$. If a point $v_0 = a_0 + i b_0$ does not belong to G , then it can be separated from G by a nonzero functional $c = x + i y$, which means $Re(c, v_0) > \sup_{v \in G} Re(c, v)$. Rewriting the scalar product in the left-hand side we obtain $(x, a_0) - (y, b_0) > \sup_{v \in G} Re(c, v)$. Note that $e^{-it} v \in G$ for all $t \in \mathbb{R}$. Substituting this for v in the right-hand side, we get $(x, a_0) - (y, b_0) > \sup_{v \in G, t \in \mathbb{R}} Re(c, e^{-it} v)$. Since $Re(c, e^{-it} v) = ((x, a) - (y, b)) \cos t + ((x, b) + (y, a)) \sin t$ and the supremum of this value over all $t \in \mathbb{R}$ is equal to

$$\sqrt{((x, a) - (y, b))^2 + ((x, b) + (y, a))^2},$$

we conclude that

$$(x, a_0) - (y, b_0) > \sup_{a+ib \in G} \sqrt{((x, a) - (y, b))^2 + ((x, b) + (y, a))^2}. \tag{8}$$

Since G is symmetric with respect to the conjugate, we have $\bar{v} \in G$ and, hence, $i\bar{v} = \mathbf{b} + i\mathbf{a} \in G$. Hence, one can interchange \mathbf{a} and \mathbf{b} in (8) and get

$$(\mathbf{x}, \mathbf{a}_0) - (\mathbf{y}, \mathbf{b}_0) > \sup_{\mathbf{a}+i\mathbf{b} \in G} \sqrt{((\mathbf{x}, \mathbf{b}) - (\mathbf{y}, \mathbf{a}))^2 + ((\mathbf{x}, \mathbf{a}) + (\mathbf{y}, \mathbf{b}))^2} \quad (9)$$

If $-(\mathbf{x}, \mathbf{a}) \cdot (\mathbf{y}, \mathbf{b}) + (\mathbf{x}, \mathbf{b}) \cdot (\mathbf{y}, \mathbf{a}) \geq 0$, then (8) yields

$$(\mathbf{x}, \mathbf{a}_0) - (\mathbf{y}, \mathbf{b}_0) > \sup_{\mathbf{a}+i\mathbf{b} \in G} \sqrt{(\mathbf{x}, \mathbf{a})^2 + (\mathbf{y}, \mathbf{b})^2 + (\mathbf{x}, \mathbf{b})^2 + (\mathbf{y}, \mathbf{a})^2}. \quad (10)$$

Otherwise, if $-(\mathbf{x}, \mathbf{a}) \cdot (\mathbf{y}, \mathbf{b}) + (\mathbf{x}, \mathbf{b}) \cdot (\mathbf{y}, \mathbf{a}) \leq 0$, then we apply (9) and arrive at the same inequality (10). Since inequality (10) is strict, we take squares of its both parts and obtain that there exists $\varepsilon > 0$ such that

$$((\mathbf{x}, \mathbf{a}_0) - (\mathbf{y}, \mathbf{b}_0))^2 > \sup_{\mathbf{a}+i\mathbf{b} \in G} ((\mathbf{x}, \mathbf{a})^2 + (\mathbf{y}, \mathbf{b})^2 + (\mathbf{x}, \mathbf{b})^2 + (\mathbf{y}, \mathbf{a})^2) + \varepsilon. \quad (11)$$

Denote by \mathbf{p} the vector from the set $\{\mathbf{x}, \mathbf{y}\}$ on which the maximum $\max_{\mathbf{p} \in \{\mathbf{x}, \mathbf{y}\}} (\mathbf{p}, \mathbf{a}_0)^2 + (\mathbf{p}, \mathbf{b}_0)^2$ is attained. Note that \mathbf{p} depends on \mathbf{c} and \mathbf{v}_0 only. Hence, for every point $\mathbf{v} = \mathbf{a} + i\mathbf{b} \in G$, we have

$$(\mathbf{p}, \mathbf{a})^2 + (\mathbf{p}, \mathbf{b})^2 \leq (\mathbf{x}, \mathbf{a})^2 + (\mathbf{x}, \mathbf{b})^2 + (\mathbf{y}, \mathbf{a})^2 + (\mathbf{y}, \mathbf{b})^2 \leq ((\mathbf{x}, \mathbf{a}_0) - (\mathbf{y}, \mathbf{b}_0))^2 - \varepsilon$$

On the other hand,

$$\begin{aligned} ((\mathbf{x}, \mathbf{a}_0) - (\mathbf{y}, \mathbf{b}_0))^2 &\leq 2(\mathbf{x}, \mathbf{a}_0)^2 + 2(\mathbf{y}, \mathbf{b}_0)^2 \\ &\leq 2((\mathbf{x}, \mathbf{a}_0)^2 + (\mathbf{x}, \mathbf{b}_0)^2 + (\mathbf{y}, \mathbf{a}_0)^2 + (\mathbf{y}, \mathbf{b}_0)^2) \\ &\leq 4((\mathbf{p}, \mathbf{a}_0)^2 + (\mathbf{p}, \mathbf{b}_0)^2). \end{aligned}$$

Thus,

$$(\mathbf{p}, \mathbf{a}_0)^2 + (\mathbf{p}, \mathbf{b}_0)^2 - \frac{\varepsilon}{4} \geq \frac{1}{4}((\mathbf{p}, \mathbf{a})^2 + (\mathbf{p}, \mathbf{b})^2)$$

and, consequently,

$$\sqrt{(\mathbf{p}, \mathbf{a}_0)^2 + (\mathbf{p}, \mathbf{b}_0)^2 - \frac{\varepsilon}{4}} \geq \frac{1}{2} \sqrt{(\mathbf{p}, \mathbf{a})^2 + (\mathbf{p}, \mathbf{b})^2}.$$

Now observe that the right-hand side of this inequality is equal to $\sup_{\mathbf{w} \in E(\mathbf{a}, \mathbf{b})} (\mathbf{p}, \mathbf{w})$ and the left-hand side is smaller than $\sup_{\mathbf{w}_0 \in E(\mathbf{a}_0, \mathbf{b}_0)} (\mathbf{p}, \mathbf{w}_0)$. Therefore, for every pair $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ such that $\mathbf{a} + i\mathbf{b} \in G$, we have

$$\sup_{\mathbf{w}_0 \in E(\mathbf{a}_0, \mathbf{b}_0)} (\mathbf{p}, \mathbf{w}_0) > \frac{1}{2} \sup_{\mathbf{w} \in E(\mathbf{a}, \mathbf{b})} (\mathbf{p}, \mathbf{w}).$$

Thus, there exists a point $\widehat{\mathbf{w}} \in E(\mathbf{a}_0, \mathbf{b}_0)$ such that $(\mathbf{p}, \widehat{\mathbf{w}}) > \frac{1}{2} \sup_{\mathbf{w} \in E(\mathbf{a}, \mathbf{b})} (\mathbf{p}, \mathbf{w})$. This holds for every point $\mathbf{a} + i\mathbf{b} \in G$ and, in particular, for each point $\mathbf{a}_k + i\mathbf{b}_k$, $k = 1, \dots, N$. Hence, the linear functional \mathbf{p} strictly separates the point $\widehat{\mathbf{w}}$ of the ellipsoid E_0 from all ellipsoids $\frac{1}{2}E_k$, i.e. from their convex hull. Therefore, $\widehat{\mathbf{w}} \notin \frac{1}{2}P$ and, hence, $E_0 \not\subset \frac{1}{2}P$. □

After Theorem 8.2 the natural question arises whether the approximation factor $1/2$ can be increased. The following theorem shows that the answer is negative.

Theorem 8.3 *The factor $q = \frac{1}{2}$ in Theorem 8.2 is sharp.*

Proof It suffices to give an example where this factor can be arbitrarily close to $1/2$. Consider the set S of pairs of vectors $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^2 \times \mathbb{R}^2$ such that \mathbf{a}, \mathbf{b} are collinear and $\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 \leq 1$. Then define $Q = \{\mathbf{a} + i\mathbf{b} : (\mathbf{a}, \mathbf{b}) \in S\}$. Thus, $Q \subset \mathbb{C}^2$.

Since each pair $(\mathbf{a}, 0)$ with $|\mathbf{a}| = 1$ belongs to S , we see that the set ReQ contains a unit disc centred at the origin. In our notation this disc can be denoted as $E(\mathbf{e}_1, \mathbf{e}_2)$, where $\mathbf{e}_1 = (1, 0)^T$ and $\mathbf{e}_2 = (0, 1)^T$. Furthermore, if $\mathbf{v} \in Q$, then $\bar{\mathbf{v}} \in Q$ and $e^{it}\mathbf{v} \in Q$ for each $t \in \mathbb{R}$. The first assertion is obvious. To prove the second one we observe that $e^{i\tau}\mathbf{v} = \mathbf{a}_\tau + i\mathbf{b}_\tau$ with $\mathbf{a}_\tau = \mathbf{a} \cos \tau - \mathbf{b} \sin \tau$ and $\mathbf{b}_\tau = \mathbf{a} \sin \tau + \mathbf{b} \cos \tau$. Clearly, \mathbf{a}_τ and \mathbf{b}_τ are collinear and $\|\mathbf{a}_\tau\|^2 + \|\mathbf{b}_\tau\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 \leq 1$. Every point of the balanced convex hull $G = \text{cob}(Q)$ has the form $\sum_{k=1}^N z_k \mathbf{u}_k = \sum_{k=1}^N |z_k| e^{i\tau_k} \mathbf{u}_k$, where $z_k = |z_k| e^{i\tau_k}$ and $\mathbf{u}_k \in Q$, $\sum_{k=1}^N |z_k| \leq 1$. Writing $t_k = |z_k|$ and $\mathbf{v}_k = e^{i\tau_k} \mathbf{u}_k$ and using that $\mathbf{v}_k \in Q$, we see that every point of G has the form $\sum_{k=1}^N t_k \mathbf{v}_k$ with all \mathbf{v}_k from Q and $\sum_{k=1}^N t_k \leq 1$.

Now let us solve Problem EE* for the set G and for the vector $t_0(\mathbf{e}_1 + i\mathbf{e}_2)$. We find the maximal positive t_0 for which this vector belongs to G . We have $t_0(\mathbf{e}_1 + i\mathbf{e}_2) = \sum_{k=1}^N t_k \mathbf{v}_k$ with $\mathbf{v}_k = \mathbf{a}_k + i\mathbf{b}_k \in Q$ and $t_k \geq 0$, $\sum_{k=1}^N t_k \leq 1$. We are going to show that $t_0 \leq \frac{1}{2}$.

Let \mathbf{a}_k be co-directed to the vector $(\cos \gamma_k, \sin \gamma_k)^T$; the vector \mathbf{b}_k has the direction $\varepsilon_k(\cos \gamma_k, \sin \gamma_k)^T$, where $\varepsilon_k \in \{1, -1\}$. Since $\|\mathbf{a}_k\|^2 + \|\mathbf{b}_k\|^2 \leq 1$, it follows that there is an angle $\delta_k = [0, \frac{\pi}{2}]$ and a number $h_k \in [0, 1]$ such that $\|\mathbf{a}_k\| = h_k \cos \delta_k$, $\|\mathbf{b}_k\| = h_k \sin \delta_k$. We have $\sum_{k=1}^N t_k \mathbf{a}_k = t_0 \mathbf{e}_1$. In the projection to the abscissa, we have $\sum_{k=1}^N t_k (\mathbf{a}_k, \mathbf{e}_1) = t_0$, and hence, $\sum_{k=1}^N t_k h_k \cos \gamma_k \cos \delta_k = t_0$. Similarly, after the projection of the equality $\sum_{k=1}^N t_k \mathbf{b}_k = t_0 \mathbf{e}_2$ to the vector \mathbf{e}_2 , we get $\sum_{k=1}^N \varepsilon_k t_k h_k \sin \gamma_k \sin \delta_k = t_0$. Taking the sum of these two equalities, we obtain $\sum_{k=1}^N t_k h_k \cos(\gamma_k - \varepsilon_k \delta_k) = 2t_0$. Since all numbers $h_k \cos(\gamma_k - \varepsilon_k \delta_k)$ do not exceed one, we conclude that $\sum_{k=1}^N t_k \geq 2t_0$ and, therefore, $t_0 \leq \frac{1}{2}$.

Hence, for the unit disc $E_0(\mathbf{e}_1, \mathbf{e}_2)$ and the set of ellipses $\{E(\mathbf{a}, \mathbf{b}) : \mathbf{a} + i\mathbf{b} \in G\}$, the solution of Problem EE* gives the approximation for Problem EE with the factor at most $\frac{1}{2}$. This is not the end yet, since Q is infinite and so G is not a balanced complex polytope. However, G can be approximated by a balanced polytope with an arbitrary precision. For the obtained balanced polytope, the approximation factor is close to $\frac{1}{2}$. Since it can be made arbitrarily close, the proof is completed. □

9 The corner cutting method

A straightforward approach to approximate the solution of Problem EE could be to replace E_0 by a sufficiently close circumscribed polygon and then to decide whether all its vertices belong to P . However, this idea turns out to be not efficient: A good approximation factor can only be obtained if this polygon has many vertices. We derive another approach based on step-by-step relaxation by cutting angles of a polygon: We iteratively search for the point of the ellipse E_0 with largest norm with respect to P and check whether the norm is less than 1. We begin with the following auxiliary problem PE (*point in ellipses*), which can be seen as a special case of Problem EE

Problem PE In the space \mathbb{R}^d there are ellipses E_1, \dots, E_N and a point \mathbf{w} . Find $\|\mathbf{w}\|_P$, where $P = \text{co}\{E_1, \dots, E_N\}$.

In particular, deciding whether $\|\mathbf{w}\|_P \leq 1$ is equivalent to a special case of Problem EE when the ellipse E_0 degenerates to a segment $[-\mathbf{w}, \mathbf{w}]$. This problem can be efficiently solved. Either precisely, by the conic programming method in subsection 9.2, or approximately by the linear programming method presented in Sect. 10.

9.1 The algorithm of corner cutting

We begin with description of the main idea and then define a routine of the algorithm.

The idea of the algorithm.

It may be assumed that E_0 is a unit circle. We construct a sequence of polygons circumscribed around E_0 as follows. The initial polygon is a square. In each iteration we cut off a corner of the polygon with the largest P -norm. So, we omit one vertex and add two new vertices. The cutting is by a line touching E_0 orthogonal to the segment connecting that vertex with the centre.

Let us denote by v_j the largest P -norm of vertices after the j th iteration (the initial square corresponds to $j = 0$). Since the norm is convex, its maximum on a polygon is attained at one of its vertices. Hence, the norm of the cut vertex is not less than the norm of each of the new vertices. Therefore, $v_{j+1} \leq v_j$, so the sequence $\{v_j\}_{j \geq 0}$ is non-increasing. If at some step we have $v_j \leq 1$, then all the vertices of the polygon after j iterations are inside P . Hence, this polygon is contained in P and, therefore, $E_0 \subset P$.

Otherwise, if $v_j > 1$, we have $E_0 \not\subset v_j \cos(\tau)P$, where τ is the smallest exterior angle of the resulting polygon. This is proved in Theorem 9.1 below. Thus, the algorithm solves Problem EE with the approximation factor $q \geq v_j \cos(\tau)$.

Comments. In each iteration we need to find the vertex with the maximal P -norm. Therefore, we need to compute a norm of each vertex by solving Problem PE. For this, we compute the norms of two new vertices in each iteration. Due to the central symmetry, one can reduce computation twice. Among two symmetric vertices we compute the norm of one of them and in each iteration we cut off both symmetric vertices.

Let τ be an arc of the unit circle connecting points α and β and denote by the $|\tau|$ its length. We assume that $|\tau| < \pi$. Denote by $\sigma = \sigma(\tau)$ the midpoint of the arc τ and

$$\mathbf{w}(\tau) = \cos(|\tau|/2)^{-1}(\mathbf{a}_0 \cos \sigma + \mathbf{b}_0 \sin \sigma).$$

where $\mathbf{a}_0, \mathbf{b}_0$ are the vectors defining the ellipse $E_0 = E(\mathbf{a}_0, \mathbf{b}_0)$. Two lines touching E_0 at the points corresponding to the ends of the arc τ meet at \mathbf{w} .

The algorithm

Initialization

Choose the maximum number of iterations J . We split the upper unit semicircle (the part of the unit circle in the upper coordinate half-plane) into arcs of equal length τ_1, τ_2 and compute the P -norms of the points $\mathbf{w}(\tau_i), i = 1, 2$. Denote by v_0 the maximum of those two norms and set $\mathcal{T} = \{\tau_1, \tau_2\}$.

Main loop—the j th iteration

We have a collection \mathcal{T} of $j + 1$ disjoint open arcs forming the upper semicircle, the P -norms of all $j + 1$ points $\mathbf{w}(\tau), \tau \in \mathcal{T}$, and the maximal norm v_{j-1} . Find an arc τ with the biggest P -norm and replace that arc by its two halves τ_1, τ_2 . Update \mathcal{T} and compute the P -norms of the points $\mathbf{w}(\tau_1)$ and $\mathbf{w}(\tau_2)$. Set v_j equal to the maximum of those two norms and of v_{j-1} .

- If $v_j \leq 1$, then $E_0 \subset P$ and STOP.
- If $v_j > \cos(\tau)^{-1}$, where τ is the minimal arc in \mathcal{T} , then $E_0 \not\subset P$ and STOP.
- If $1 < v_j \leq \cos(\tau)^{-1}$ and $j = J$, then $E_0 \not\subset \cos(\tau)P$.
- Otherwise go to the next iteration.

Theorem 9.1 *The corner cutting algorithm after j iterations solves Problem EE with the approximation factor $q \geq v_j \cos(\tau)$, where τ is the minimal arc in \mathcal{T} .*

Proof Let τ be the smallest arc after j iterations. Suppose this arc appears after the k th iteration, $k \leq j$. Then, its mother arc (let us call it 2τ) had the biggest value $\|\mathbf{w}(\cdot)\|_P$ among all arcs in the k th iteration. This means that $\|\mathbf{w}(2\tau)\|_P = v_k$. Since the sequence $\{v_i\}_{i \geq 0}$ is non-increasing, we have $v_k \geq v_j$. The point $\mathbf{x} = \cos(\tau)\mathbf{w}(2\tau)$ lies on E_0 . It does not belong to P precisely when $\|\mathbf{x}\|_P > 1$, i.e. when $\mathbf{w}(2\tau) > \cos(\tau)^{-1}$. Thus, if $v_j > \cos(\tau)^{-1}$, then $\mathbf{w}(2\tau) = v_k > \cos(\tau)^{-1}$ and, hence, $\mathbf{x} \notin P$. Therefore, the inequality $v_j > \cos(\tau)^{-1}$ implies that E_0 is not contained in P . □

The length of each arc has the form $2^{-s}\pi$, where s is the number of double divisions to arrive at that arc. We call this number the *level* of the arc. So, the original arcs of length $\pi/2$ are those of level one.

The complexity of the corner cutting algorithm

To perform j iterations one needs to solve Problem PE for $j + 2$ points $\mathbf{w}(\cdot)$. So, the complexity of the algorithm is defined by the complexity of solution of Problem PE. Below, in Sects. 9.2 and 10 we derive two methods of its solution, based on different ideas and compare them by numerical experiments. The approximation factor is $\cos(\tau) = \cos(2^{-s}\pi) = 1 - 2^{-2s-1}\pi^2 + O(2^{-4s})$, where s is the maximal level of the intervals after j iterations. Already for $s = 2$ (after one iteration) the approximation factor is $q = \cos(\frac{\pi}{4}) = 1/\sqrt{2}$, which is better than in the complex polytope method, where $q = \frac{1}{2}$. For $s = 3$ (after at most three iterations), we have $q = \cos(\frac{\pi}{8}) = 0.923\dots$, for $s = 5$, we have $q = 0.995\dots$, for $s = 10$, we have $q > 1 - 10^{-5}$. In the worst case reaching the level s requires $j = 2^s - 1$ iterations. However, in practice it is much faster. Numerical experiments show that j usually does not exceed $s + 2$.

In each iteration of the corner cutting algorithm we need to find the P -norm of the newly appeared vertices of the polygon. This means that we solve Problem PE for those vertices. The way of arriving at the solution actually defines the efficiency of the whole algorithm. We present two different methods and compare them.

9.2 Solving Problem PE via conic programming

The norm $\|\mathbf{w}\|_P$ is equal to the minimal $r \in \mathbb{R}$ such that $\mathbf{w} \in rP$, i.e. the minimal possible sum of non-negative numbers $r_1, \dots, r_j \in \mathbb{R}$ such that $\mathbf{w} = \sum_{j=1}^N r_j E_j$. Thus, we obtain

$$\begin{cases} r = \min \sum_{j=1}^N r_j & \text{subject to} \\ \tau_j \in [0, 2\pi), & j = 1, \dots, N, \\ r_j \geq 0, & j = 1, \dots, N. \\ \sum_{j=1}^N r_j \mathbf{a}_j \cos \tau_j + r_j \mathbf{b}_j \sin \tau_j = \mathbf{w}, \end{cases} \tag{12}$$

Changing variables $c_j = r_j \cos \tau_j, s_j = r_j \sin \tau_j$ we obtain the conic programming problem

$$\begin{cases} \text{minimize } \sum_{j=1}^N r_j & \text{subject to} \\ \sum_{j=1}^N c_j \mathbf{a}_j + s_j \mathbf{b}_j = \mathbf{w}, \\ (c_j^2 + s_j^2)^{1/2} \leq r_j, & j = 1, \dots, N. \end{cases} \tag{13}$$

with $3N$ variables $r_j, c_j, s_j \in \mathbb{R}$ and $N(d + 2)$ constraints. Among these constrains, there are $N(d + 1)$ linear and only N quadratic ones, but the latter actually define the complexity of this problem. The problem is solved by conic programming. This can be done efficiently for dimensions $d \leq 20$ and number of ellipsoids $N \leq 1000$.

The value r of the problem (12) is equal to the norm $\|\mathbf{w}\|_P$. In particular, $\mathbf{w} \in P$ if and only if $r \leq 1$.

In the next section we introduce the second approach, when the conic programming (13) problem is approximated with a linear programming one with precision that increases exponentially with the number of extra variables.

10 The projection method

The corner cutting method makes use of a polygonal approximation of the ellipse E_0 . Can we go further and approximate all the N ellipses E_1, \dots, E_N and thus approximate Problem PE with a linear programming (LP) problem? In principle, this is possible, but very inefficient. Cutting corners of N polygons is expensive and slow. If we do not involve cutting but just approximate each ellipse by a polygon, the situation will be still worse due to a large total number of vertices of all polygons. Nevertheless, each approximating polygon can be built much cheaper if we present it as a projection of a higher dimensional polyhedron. This technique was suggested by Ben-Tal and Nemirovski [2] for approximating quadratic problems by LP problems. See also [8] for generalizations to other classes of functions. We briefly describe this method (with slight modifications) and then apply it to Problem PE. Note that, in contrast to the conic programming, here we obtain only an approximate solution of Problem PE. However, this is not a restriction since the corner cutting algorithm also gives only an approximate solution for Problem EE. If q_1 and q_2 are approximation factors of those two problems, then the resulting approximation factor is q_1q_2 . If $q_i = 1 - \varepsilon_i$ with a small $\varepsilon_i, i = 1, 2$, then $q_1q_2 > 1 - \varepsilon_1 - \varepsilon_2$.

10.1 A fast approximation of ellipses

The projection method realizes a polygonal approximation of ellipses by solving a certain LP problem and the precision of this approximation increases exponentially in the LP problem input. This is done by an iterative algorithm, whose main loop is a doubling of a convex figure.

Doubling of a figure

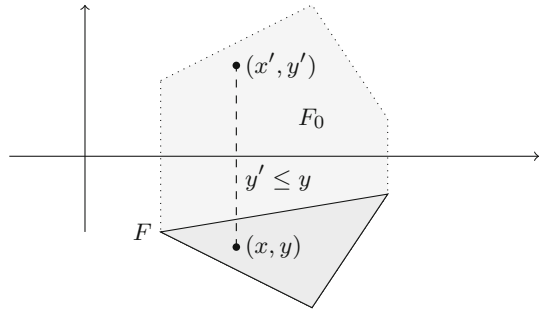
Consider an arbitrary figure $F \subset \mathbb{R}^2$ located in the lower half-space of the Cartesian plane. Then the set

$$F_0 = \{(x', y')^T : x' = x, |y'| \leq -y, (x, y)^T \in F\} \tag{14}$$

is the convex hull of F with its reflection about the abscissa, see Fig. 1. Indeed, each point $A = (x, y)^T \in F$ produces a vertical segment $\{(x, y') : y' \in [y, -y]\}$ which connects A with its reflection A' about the abscissa. Those segments fill the set F_0 .

In the same way one can double a figure F about an arbitrary line passing through the origin provided F lies on one side with respect to this line. Let a line ℓ_α be defined by the equation $y = x \tan \alpha$; it makes the angle $\alpha \in [0, \pi]$ with the abscissa. After the clockwise rotation by the angle α the line ℓ_α becomes the abscissa and F becomes

Fig. 1 Set F and the convex hull with its reflection at the abscissa
 $F_0 = \text{co}\{F, F'\}$



a figure F' located in the lower half-plane. Since this rotation is defined by the matrix $R_\alpha = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}$, it follows from formula (14) that the figure F_α , the convex hull of F with its reflection about the line ℓ_α , consists of points (x_1, y_1) satisfying the following system of inequalities:

$$\begin{cases} x_1 \cos \alpha + y_1 \sin \alpha = x \cos \alpha + y \sin \alpha \\ |-x_1 \sin \alpha + y_1 \cos \alpha| \leq x \sin \alpha - y \cos \alpha \\ (x, y)^T \in F \end{cases} \quad (15)$$

Construction of a regular 2^n -gon

Now we describe the algorithm of recursive doubling of a polygon.

We take an arbitrary radius $r > 0$, denote $\alpha_m = 2^{-m}\pi, m \geq 0$, and consider an isosceles triangle AOB , where $A = (r, 0)^T, B = (r \cos \alpha_n, r \sin \alpha_n)^T$, and O is the origin. Double this triangle about the line $\ell_{\alpha_n} = OB$, then double the obtained quadrilateral about $\ell_{\alpha_{n-1}}$ (the lateral side different from OA), then about $\ell_{\alpha_{n-2}}$, etc.. After n doublings (the last one is about ℓ_1 , which is abscissa) we get the regular 2^n -gon inscribed in the circle of radius r . We denote this polygon by rT_n . Thus, T_n is the 2^n -gon inscribed in the unit circle. Note that the initial triangle AOB is defined by the system of linear inequalities

$$\begin{aligned} 0 \leq y \leq x \tan \alpha_n, \quad \text{and} \\ x + y \tan \alpha_{n+1} \leq r. \end{aligned}$$

Thus, we obtain the following description of the set rT_n , which is a regular 2^n -gon inscribed in the circle of radius r :

$$\begin{aligned} rT_n &= (x_{2n+1}, x_{2n+2})^T \\ \left\{ \begin{aligned} 0 \leq x_2 \leq x_1 \tan \alpha_{n-1} \\ x_1 + x_2 \tan \alpha_n \leq r \\ \text{for } k = 1, \dots, n : \\ x_{2k+1} \cos \alpha_{n-k} + x_{2k+2} \sin \alpha_{n-k} &= x_{2k-1} \cos \alpha_{n-k} + x_{2k} \sin \alpha_{n-k} \\ |-x_{2k+1} \sin \alpha_{n-k} + x_{2k+2} \cos \alpha_{n-k}| &\leq x_{2k-1} \sin \alpha_{n-k} - x_{2k} \cos \alpha_{n-k} \end{aligned} \right. \quad (16) \end{aligned}$$

in the variables $k = 1, \dots, n, r, x_1, \dots, x_{2n+2}$. The inequality with modulus $|a| \leq b$ is replaced by the system $a \leq b, -a \leq b$. The system (16) consists of $3n + 3$ linear constraints (equations and inequalities) with $2n + 3$ variables. For all vectors $X = (x_1, \dots, x_{2n+2})^T$ satisfying system (16), the vector composed by the two last components $(x_{2n+1}, x_{2n+2})^T$ fills the regular 2^n -gon. So, this 2^n -gon is a projection of a $(2n + 2)$ -dimensional polyhedron to the plane. This polyhedron has $3n + 3$ facets.

Construction of an affine-regular 2^n -gon inscribed in an ellipse

For an arbitrary ellipse $E(\mathbf{a}, \mathbf{b})$, the point $x_{2n+1}\mathbf{a} + x_{2n+2}\mathbf{b}$ runs over an affine-regular 2^n -gon inscribed in the ellipse $rE(\mathbf{a}, \mathbf{b})$ as the vector $X = (x_1, \dots, x_{2n+2})$ runs over the set of solutions of the linear system (16) with this value of r .

10.2 Solving Problem PE by the fast polygonal approximation

We approximate all ellipses $E_j = E(\mathbf{a}_j, \mathbf{b}_j), j = 1, \dots, N$ by polygons and then decide whether $\mathbf{w} \in P$ with some approximation factor.

We fix a natural n and non-negative numbers $r^{(1)}, \dots, r^{(N)}$ such that $\sum_{j=1}^N r^{(j)} = 1$. For each j , we consider the affine-regular polytope $r_j T_n^{(j)} = x_{2n+1}^{(j)} \mathbf{a}_j + x_{2n}^{(j)} \mathbf{b}_j$ inscribed in E_j , where $(r_j, X^{(j)}) = l(r_j, x_1^{(j)}, \dots, x_{2n+1}^{(j)}, x_{2n+2}^{(j)})^T$ is a feasible vector for the linear system (16). If $\mathbf{w} \in r^{(1)} T_n^{(1)} + \dots + r^{(N)} T_n^{(N)}$, then $\mathbf{w} \in r^{(1)} E_1 + \dots + r^{(N)} E_N$. Therefore, $\mathbf{w} \in P$ whenever there exist numbers $r^{(j)} \geq 0$ such that $\sum_{j=1}^N r^{(j)} = 1$ and $\mathbf{w} \in r^{(1)} T_n^{(1)} + \dots + r^{(N)} T_n^{(N)}$. Hence, the assertion $\mathbf{w} \in P$ is decided by the following LP problem:

$$\left\{ \begin{array}{l} \sum_{j=1}^N r^{(j)} \rightarrow \min, r^{(j)} \geq 0 \\ 0 \leq x_2^{(j)} \leq x_1^{(j)} \tan \alpha_{n-1}, x_1^{(j)} + x_2^{(j)} \tan \alpha_n \leq r^{(j)} \\ x_{2k+1}^{(j)} \cos \alpha_{n-k} + x_{2k+2}^{(j)} \sin \alpha_{n-k} = x_{2k-1}^{(j)} \cos \alpha_{n-k} + x_{2k}^{(j)} \sin \alpha_{n-k} \\ \left| -x_{2k+1}^{(j)} \sin \alpha_{n-k} + x_{2k+2}^{(j)} \cos \alpha_{n-k} \right| \leq x_{2k-1}^{(j)} \sin \alpha_{n-k} - x_{2k}^{(j)} \cos \alpha_{n-k} \\ \mathbf{w} = \sum_{j=1}^N x_{2n+1}^{(j)} \mathbf{a}_j + x_{2n+2}^{(j)} \mathbf{b}_j \end{array} \right. \quad (17)$$

in the variables $k = 1, \dots, n, r^{(j)}, x_s^{(j)}, j = 1, \dots, N, s = 1, \dots, 2n + 2$. Let us remember that $\alpha_m = 2^{-m}\pi$. The value of this problem $r = \sum_{j=1}^N r^{(j)}$ is the minimal number such that \mathbf{w} belongs to the set rP_n , where $P_n = \text{co} \{T_n^{(1)}, \dots, T_n^{(N)}\}$. In other words, $r = \|\mathbf{w}\|_{P_n}$. In particular, $\mathbf{w} \in P_n$ precisely when $r \leq 1$.

The LP problem (17) has $(2n + 3)N$ variables $r^{(j)}, x_s^{(j)}$ and $(3n + 4)N + d + 1$ linear constraints (equations and inequalities). Note that the matrix of this problem possesses only $(12n + 2d + 7)N + d$ nonzero coefficients, i.e. the total number of nonzero coefficients is linear in the size of the matrix. On the other hand, the product of the number of variables times the number of constraints exceeds $6n^2 N^2 + 2Nd$. Thus, this problem is very sparse.

Table 1 The partial approximation factor q_1 for Problem PE for small n (values rounded)

n	3	4	5	6	7	8
$\cos(2^{-n}\pi)$	0.9238	0.9807	0.9951	0.9987	0.9996	0.9999

Since $P_n \subset P$, it follows that $\mathbf{w} \in P$ whenever $r \leq 1$. In fact, problem (17) provides an approximate solution to Problem PE with the factor $q = \cos(2^{-n}\pi)$.

Theorem 10.1 *If r is the value of the LP problem (17), then for every $\mathbf{w} \in \mathbb{R}^d$ we have $r \cos(2^{-n}\pi) \leq \|\mathbf{w}\|_P \leq r$.*

Proof Since the ratio between the radii of the inscribed and the circumscribed circles of a regular 2^n -gon is equal to $q = \cos(2^{-n}\pi)$, we see that $E_j \subset qT_n^{(j)}$ for each j . Consequently, $qP \subset P_n$ and $\|\mathbf{w}\|_P \geq q \|\mathbf{w}\|_{P_n}$, from which the theorem follows. \square

Corollary 10.2 *If $r \leq 1$, then $\mathbf{w} \in P$, otherwise $\mathbf{w} \notin \cos(2^{-n}\pi)P$.*

Since $\cos(2^{-n}\pi) = 1 - 2^{-2n-1}\pi^2 + O(2^{-4n})$, we see that already for small values of n we obtain a very sharp estimate. The rate of approximation for $n \leq 8$ is given in Table 1. For $n = 12$ we have $q > 1 - 10^{-6}$; for $n = 17$ we have $q > 1 - 10^{-9}$.

11 Numerical results

In this section we demonstrate practical implementations of our methods of finding the convex hulls of ellipses.¹

We obtain numerical results and compare them for the following methods presented in this paper:

- Complex polytope method (Sect. 8)
- Corner cutting method (Sect. 9)
- Projection method (Sect. 10)
- Mixed method

The *Mixed method* is a combination of the complex polytope method and the projection method. The former is the fastest algorithms of all three, the latter is the most accurate. The mixed method accepts an additional parameter *bound* describing the range of values one is interested in. Whenever the complex polytope method determined that

¹ We use the following solvers: *Matlab's* `linprog` and *Gurobi* for the linear programming (LP) problems and *Gurobi* for the quadratic programming (QP) problems. *SeDuMi Gurobi* is a commercial solver, but a free academic license can be obtained at gurobi.com *SeDuMi* is free and can be downloaded at github.com/sqjp/SeDuMi. The GitHub version is a maintained fork of the original project, whereas the original host does not seem to maintain SeDuMi any more. For the tests we used a PC with an AMD Ryzen 3600, 6 cores (5 cores used), 3.6 GHz, 64 GB RAM, Windows 10 build 1809 Matlab R2020a, Gurobi solver 9.0.2 from May 2019, SeDuMi solver 1.32 from July 2013, *toolboxes* v1.2 from June 2021, TTEST v0.9 from June 2021. The algorithms are implemented in Matlab and included in the *toolboxes* [24]. The scripts to generate and evaluate the data can be downloaded from tommsch.com/science.php All software is thoroughly tested using the TTEST framework [25].

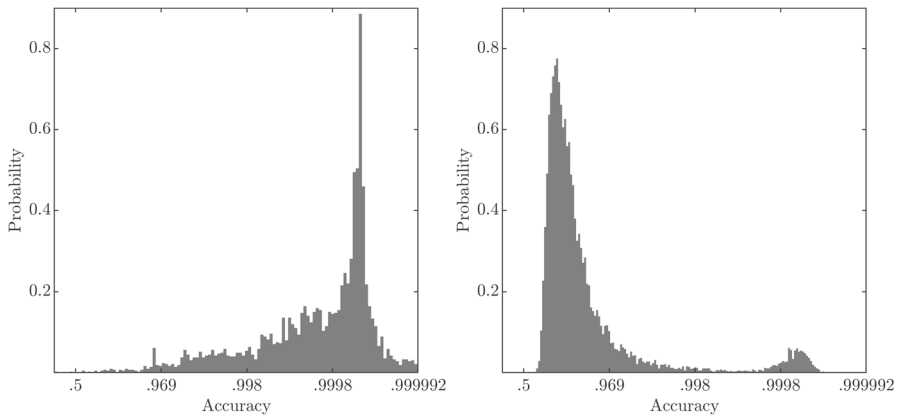


Fig. 2 Estimated probability density function of the approximation factor of the complex polytope method for two different data sets. The left pictures data set is Dataset (A), the right pictures data set is Dataset (B). The x axis is an inverted, and shifted by 1, logarithmic scale

the norm is inside or outside of the range of interest, the exact algorithm is not started, and thus the computation is sped up. For example, for the application of computing the joint spectral radius using the Invariant polytope algorithm, one is only interested whether an ellipse lies inside or outside of the convex hull of the elliptic polytope. Now, whenever the complex polytope method concludes that an ellipse lies inside or outside, one can already abort the computation.

The various implemented methods are optimized to a different degree, and thus, timings cannot be compared well.

To obtain quantitative measures of how the methods differ, we generated two test sets of random ellipses and elliptic polytopes.

(Dataset A) The first set contains ellipses and elliptic polytopes whose ellipses have normal distributed real and imaginary part. Dataset (A) consists of 365 elliptic polytopes in dimension 3 to 25 and the norm is computed approximately for 12 ellipses per elliptic polytope.

(Dataset B) The second set is generated by the Invariant polytope algorithm, where we stored the intermediate occurring ellipses and elliptic polytopes for some random sets of input matrices with complex leading eigenvalue. Dataset (B) consists of 119 elliptic polytopes in dimensions 2 to 12 and the norm is computed for 100 ellipses per elliptic polytopes.

The measured runtime of the algorithms with respect to the chosen accuracy and number of vertices can be seen in Fig. 9.

11.1 Behaviour of the complex polytope method

Although the theoretical approximation factor of this method is $1/2$, in numerical experiments it turns out that the average approximation factor is mostly larger than $1/\sqrt{2}$. In small dimensions, $d = 2, 3$, the approximation factor is even close to 1 in a lot of cases, see Fig. 2 for the (estimated) probability density function of this method's

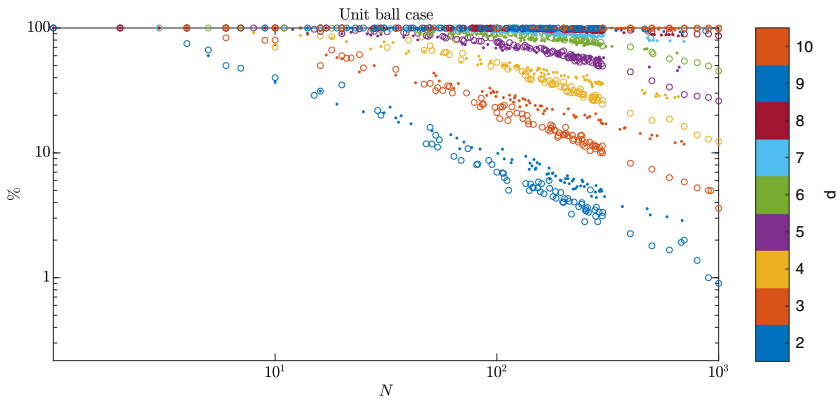


Fig. 3 Fraction of points or ellipses which belong to the convex hull of randomly selected points or ellipses which are uniformly distributed in the unit ball or have uniformly distributed real and imaginary part in the unit ball, respectively. Since the computational time increases significantly with the number of points or ellipses, for sets with more than 300 points or ellipses less examples were conducted. In the plot one can see the relative fraction of points or ellipses belonging to the convex hull, coloured with respect to the dimension. The point examples are plotted with a \cdot symbol, the ellipse examples are plotted with a \circ symbol

approximation factors.

Note that the numerical accuracy of the QP solver is approximately 10^{-5} and thus, the maximal accuracy which can be reached is approximately 0.99999, which is quite exactly the position of the rightmost peaks in Fig. 2.

11.2 Behaviour of the corner cutting method

The corner cutting method is, like the complex polytope method, a QP problem, and thus, the absolute error of the solution returned by our numerical solvers is in the range of 10^{-5} . For the corner cutting method, this accuracy is on average obtained after 10 to 12 iterations in the generic case, as our experiments show. Apart from the chosen accuracy, the runtime of the algorithm mostly depends on, firstly, the geometry of the problem and, secondly, on the number of vertices of the elliptic polytope. The dimension of the problem only has a minor influence on the runtime.

11.3 Behaviour of the projection method

The absolute error of the LP solvers is roughly 10^{-9} , which is magnitudes higher than for the QP solver, which is typically around 10^{-5} . Solely due to this fact, the projection method is the most accurate method of all the described methods.

For the projection method one could increase the number of vertices of the polytopes approximating the ellipses of the elliptic polytope until the norm is computed up to the desired accuracy, similarly as in the corner cutting method. Unfortunately, this hinders the use of warm-starting the LP problem since this alters the underlying

LP. Therefore, in our implementation we choose the approximation factor q_1 corresponding to Problem EE* to be of the same magnitude as the approximation factor q_2 corresponding to Problem EE, and such that $q_1 q_2 \simeq q$, where q is the chosen accuracy.

12 Applications

12.1 Number of extremal vertices

We address the question of the expected number of vertices in the convex hull of random ellipses. This issue is important for both of the above applications since it shows the growth of the number of ellipses with respect to the number of the iterations of the algorithms.

The corresponding problem for the convex hull of random points originated with the famous question of Sylvester [30]. The answer highly depends on the domain on which the points are sampled and on the dimension. Various lower and upper bounds on the asymptotically expected number of points in the convex hull are known, see [4, 18] and references therein. It would be extremely interesting to come up with similar theoretical estimates for convex hulls of ellipses. Here we compare the two cases solely numerically. There is no canonical analogue between points sampled from some domain and ellipses sampled from some domain, since the ellipses are determined by two parameters instead of one. We introduce several numerical results with various samplings.

Uniformly sampled ellipsoids in the unit ball

We are given ellipses whose real and imaginary parts are sampled uniformly from the unit ball and points uniformly sampled from the unit ball as well. The number of vertices and ellipses of their corresponding convex hull is computed and plotted in Fig. 3. Experiments are made for dimensions 2 to 10 and number of points and ellipses 1 to 1000. Interestingly, the two cases differ greatly. Whereas for dimensions 2 to 5 the fraction of ellipses belonging to the convex hull is less than for the point counterpart, the situation is reversed from dimension 7 upwards.

Uniformly sampled ellipsoids in the unit cube

Interestingly, when the points or the real and imaginary parts of the ellipses are sampled uniformly from a unit-cube, the behaviour between the point case and the ellipses is very similar, at least for small dimensions, as can be seen in Fig. 4.

Gaussian sampled like ellipsoids

Also for points and ellipses with real and imaginary part sampled from a d -dimensional normal distribution, the behaviour is similar. See Fig. 5 for a visualization of the obtained numerical results.

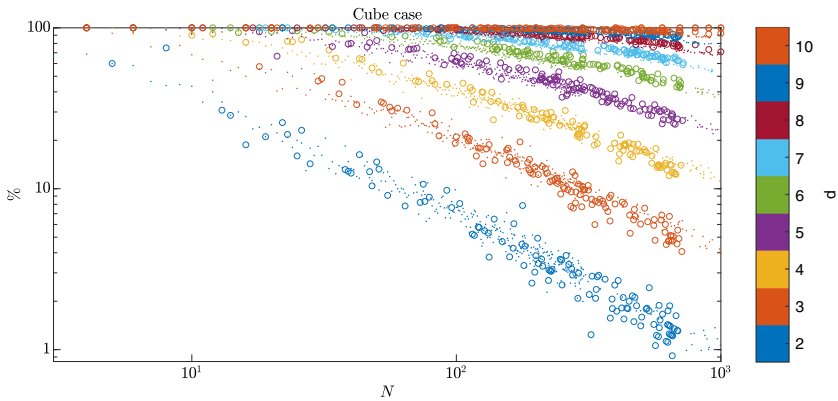


Fig. 4 Fraction of points or ellipses which belong to the convex hull of randomly selected points or ellipses which are uniformly distributed in the unit cube or have uniformly distributed real and imaginary part in the unit cube, respectively

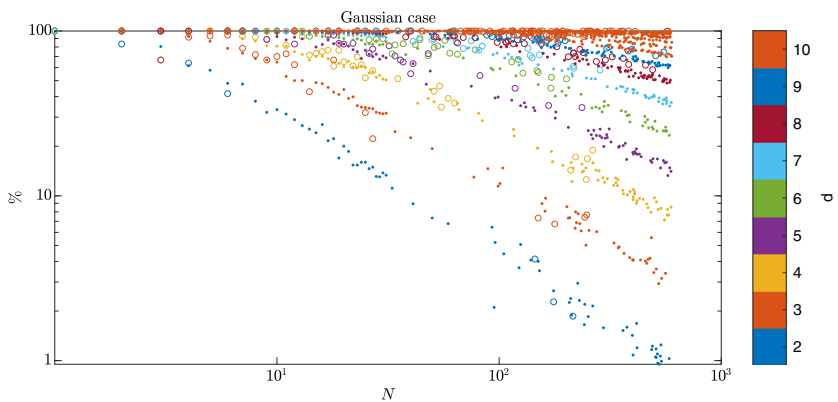


Fig. 5 Fraction of points or ellipses which belong to the convex hull of randomly selected points or ellipses which are normal distributed or have normal distributed real and imaginary part, respectively

12.2 Lyapunov function for a discrete time linear system

Given a linear system defined by a $d \times d$ matrix A with a complex leading eigenvalue, which is supposed to be unique and simple (up to complex conjugates); We need to construct a norm $\|\cdot\|$ in \mathbb{R}^d such that $\|Ax\| \leq \rho(A)\|x\|$ for all $x \in \mathbb{R}^d$. This is the same as constructing a symmetric convex body $P \subset \mathbb{R}^d$ for which $AP \subset \rho(A)P$. It is obtained as an elliptic polytope by an iteration method, see Sect. 1, Application 2. In Fig. 6 the time needed to compute the invariant elliptic polytope P is plotted against the dimension. The colour indicates the number of vertices of the set V .

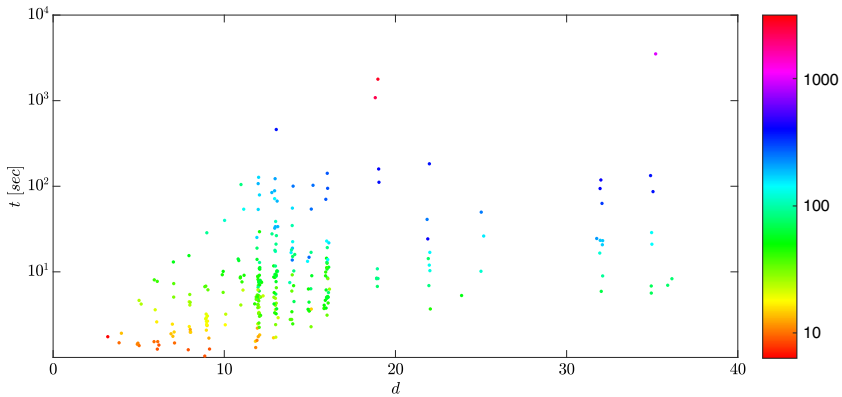
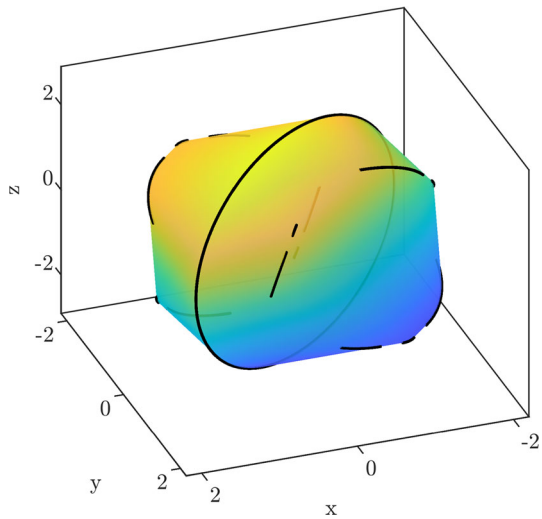


Fig. 6 The computation time for evaluating an invariant elliptic polytope P for a given matrix A with complex leading eigenvalue such that $AP \subset \rho(A)P$ holds. The colour indicates the number of vertices of the polytope

Fig. 7 Invariant elliptic polytope for the matrices $\begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}$. The colour indicates the z -value of the convex hull. The black lines are drawn where the ellipses touch the boundary of the convex hull



12.3 Spectral characteristics of matrices

As already noted, one application is the computation of the joint spectral radius, for example used for the computation of the Lyapunov function of a discrete time linear system.

Example 12.1 Given $A = \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}$; The invariant polytope algorithm computes $\rho := JSR(\{A, B\}) = \rho(B) = \sqrt{2}$ and that, with the eigen-ellipse $E_0 = E(v_0) = E\left(\begin{bmatrix} 2 \\ 1-\sqrt{-7} \end{bmatrix}\right)$, i.e. $BE_0 = E_0$, the elliptic polytope $P = co\{E_0, AE_0, A^2E_0, BAE_0, A^3E_0, BA^2E_0, B^2AE_0, BA^3E_0, B^2A^2E_0, B^2A^3E_0\}$ fulfils $AP \subset \rho P$ and $BP \subset \rho P$, see Fig. 7 for a plot of P .

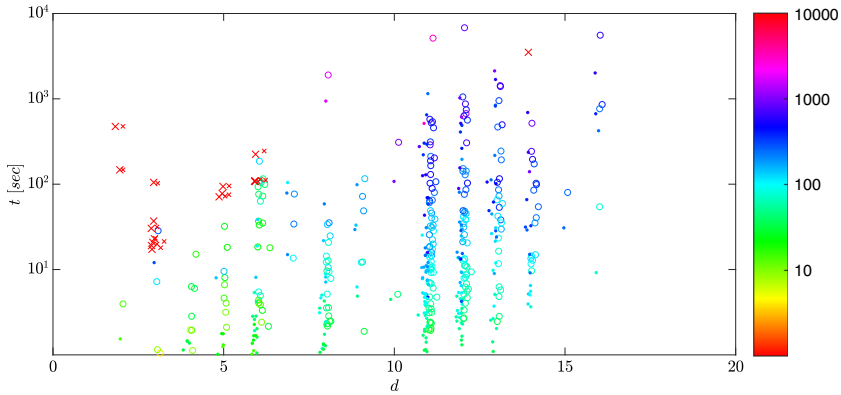


Fig. 8 The time of computation of the joint spectral radius for a pair of matrices $A_1, A_2 \in \mathbb{R}^d$, whose spectrum maximizing product has complex leading eigenvalue. The colour indicates the number of vertices of the polytope. On the x-axis we have the dimension of the generated example, the y-axis shows the time needed to compute an invariant polytope. The x-values are slightly distorted for better readability. Similar to the case of random matrices with real leading eigenvalue, it seems that the existence of a spectrum maximizing product with finite length is generic. The results obtained using the complex-polytope method are marked with a \cdot symbol, the results obtained using the projection method are marked with a \circ symbol. Examples where the algorithm could not find an invariant polytope are marked in both cases with a red \times symbol. We suspect the reason why the Invariant polytope algorithm may not terminate within reasonable time for certain examples is a long spectrum maximizing product for the set under test

In Fig. 6 the time needed to compute an invariant elliptic polytope P is plotted against the dimension. Currently, the invariant polytope algorithm can compute the joint spectral radius of pairs of matrices of up to dimension 17, for which, in median, the computational time is 1000 s and the invariant polytope consists of 1400 ellipses. It seems that the time increases sub-exponentially with the dimension and depends linearly on the number of ellipses of the invariant elliptic polytope. In other words, the number of ellipses of the elliptic polytope increases sub-exponentially with the dimension.

Now we analyse the performance of the Invariant polytope algorithm for computation of the joint spectral radius of a set of matrices. The elliptic polytopes are applied in the case when the spectrum maximizing product has a complex leading eigenvalue. In the construction of the invariant elliptic polytopes we can use each of our methods. The numerical tests show that the projection method always performs better than the corner cutting method and that the mixed method always performs better than the projection method. Thus, only two significant algorithms remain, the complex polytope method and the projection method. We are comparing them.

In Fig. 8 and 9 the results of our experiments are plotted.

Remark 12.2 In practice it occurs rather seldom that a spectrum maximizing product of a set of matrices possesses a complex leading eigenvalue and whose length is greater than one. One such example is given in the Appendix, Example A.3.

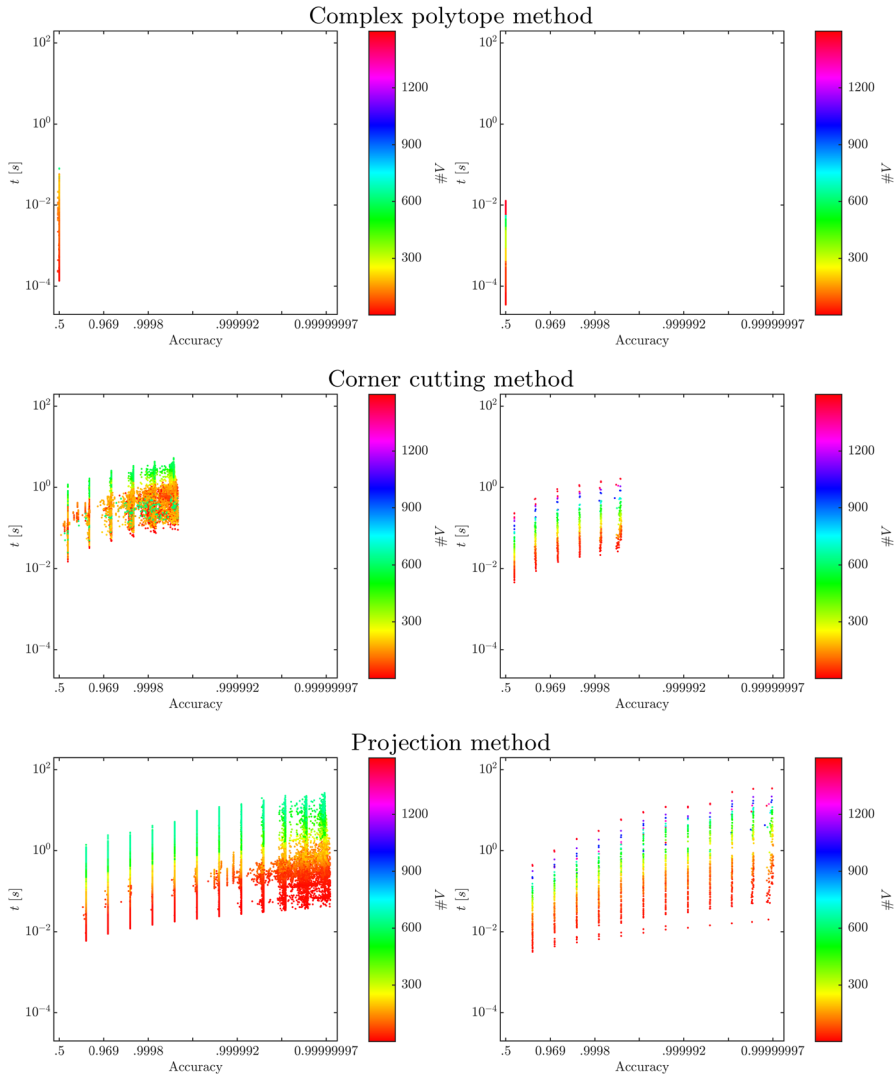


Fig. 9 Runtime t in seconds of the the methods *complex polytope*, *corner cutting* and *projection*. On the x -axis the theoretical minimal accuracy on a logarithmic scale is printed, on the y -axis the time the algorithm needed, also on a logarithmic scale. The true, obtained, accuracy is especially for the complex polytope method much higher. The colour indicates the number of vertices of the elliptic polytope. The dimension of the problem is not plotted, since it turned out to have only a very minor influence on the runtime. All algorithms were assessed using the same data set, the corner cutting method and the projection method were tested with different approximation factors. The left column is for data arising in the Invariant polytope algorithm. The right column is for data of ellipses and elliptic polytopes with normal distributed real and imaginary part. One can see clearly, that the complex polytope method is the most efficient algorithm when one compares the time the algorithm needs with its accuracy. This is even more true under the viewpoint that the complex polytope method on average yields an accuracy of 0.7071. Comparing the corner cutting method and the projection method, one sees that the latter clearly outperforms the former consistently

Acknowledgements The authors are grateful for the referee's valuable and profound comments.

Author Contributions All authors contributed in equal parts to this paper.

Funding Thomas Meistrík is funded by FWF, grant P 33352.

Data availability The main algorithms are available at [gitlab.com/tommsch](https://github.com/tommsch), all supplementary data, including the scripts to generate the data is available at tommsch.com/science.php.

Declarations

Conflict of interest The authors have no competing interests.

Ethics approval Not applicable.

A Appendix

Theorem (Theorem 5.4) *For each $N \geq 2$ there exists a polyhedron in \mathbb{R}^N with less than $2N$ facets and a positive semidefinite quadratic form of rank two on that polyhedron which has at least 2^{N-2} points of local maxima with different values of the function.*

We begin with the following technical fact. Let us have a vector $\mathbf{a} \in \mathbb{R}^2$ and a line ℓ on \mathbb{R}^2 which is not parallel to \mathbf{a} . An *affine symmetry* about ℓ along \mathbf{a} is an affine transform that, for each $\mathbf{x} \in \ell$ and $t \in \mathbb{R}$, maps the point $\mathbf{x} + t\mathbf{a}$ to $\mathbf{x} - t\mathbf{a}$. If $\mathbf{a} \perp \ell$, then this is the usual (orthogonal) symmetry.

Lemma A.1 *Let O be an arbitrary point on the side of a convex polygon different from its midpoint. Then there exists an affine symmetry about this side arbitrarily close to an orthogonal symmetry such that the distances from O to the images of the vertices of this polygon are all different.*

Proof Let O be the origin and one of the basis vectors along that side. Then, the matrix of an arbitrary affine symmetry is $S = \begin{pmatrix} 1 & a \\ 0 & -1 \end{pmatrix}$, with $a \in \mathbb{R}$ arbitrary. If the images $A\mathbf{x}$ and $A\mathbf{y}$ of two vertices $\mathbf{x} \neq \mathbf{y}$ are equidistant from O , then $A(\mathbf{x} + \mathbf{y})$ and $A(\mathbf{x} - \mathbf{y})$ are orthogonal and, hence, $(\mathbf{x} - \mathbf{y})A^T A(\mathbf{x} + \mathbf{y}) = 0$. This quadratic equation in a has at most two solutions. Hence, there exists only a finite number of values of a for which some of the images of the vertices are equidistant from O . \square

Proposition A.2 *For every $n \geq 2$ and $\varepsilon > 0$, there exists a polyhedron Q_n in \mathbb{R}^{n+2} with at most $2n + 3$ facets whose orthogonal projection to some two-dimensional plane is a 2^n -gon such that: (1) Its distance (in the Hausdorff metric) to a regular 2^n -gon centred at the origin is less than ε . (2) The distances from its 2^n vertices to the origin are all different.*

Proof Applying the construction (16) for $r = 1$, we obtain a polyhedron that consists of points $(x_1, \dots, x_{2n+2})^T \in \mathbb{R}^{2n+2}$ satisfying the system (16). That system contains n linear equations and $2n + 3$ linear inequalities. Hence, it defines an $(n + 2)$ -dimensional polyhedron with at most $2n + 3$ facets. Its projection to the plane (x_{2n+1}, x_{2n+2}) is

a regular 2^n -gon. Now, in each iteration $j = 1, \dots, n$ of the construction (16), we replace the symmetry about the line $\ell_{\alpha_{n-j+1}}$ by a close affine symmetry about the same line. Invoking Lemma A.1 we can choose this symmetry so that the resulting polygon has all its vertices on different distances from the origin. Hence, the polygon obtained after the last iteration also possesses this property. \square

Proof of Theorem 5.4. After applying Proposition A.2 for $n = N - 2$, we obtain a polyhedron $Q_{N-2} \subset \mathbb{R}^N$ whose two-dimensional projection to the plane (x_{2N-3}, x_{2N-2}) is a 2^{N-2} -gon close to a regular 2^{N-2} -gon. Then, for the quadratic form $x_{2N-3}^2 + x_{2N-2}^2$, each vertex of this polygon is a local maximum and all the values in those points are different. \square

The following example is for the curious reader, who asks herself whether there exists a set of matrices, for which the spectral maximizing product has length greater than one and a complex leading eigenvalue.

Example A.3 For $\alpha, \beta \in (-\pi/2, \pi/2)$, $\alpha \neq \beta$, the set $\{T_0, T_1\}$,

$$T_0 = \begin{pmatrix} 0 & 0 & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ \cos \alpha & \sin \alpha & 0 \end{pmatrix}, \quad T_1 = \begin{pmatrix} 0 & -\sin \beta & \cos \beta \\ 0 & \cos \beta & \sin \beta \\ 0 & 0 & 0 \end{pmatrix},$$

has $T_0 T_1$ as spectrum maximizing product, i.e. up to permutations and powers the normalized spectral radius of all other products of matrices T_0 and T_1 is strictly less than $\rho(T_0 T_1)^{1/2} = 1$ (Fig. 9).

References

1. Barabanov, N.E.: Lyapunov indicator for discrete inclusions, I–III. *Autom. Remote Control* **49**(2), 152–157 (1988)
2. Ben-Tal, A., Nemirovski, A.: On polyhedral approximations of the second-order cone. *Math. Oper. Res.* **26**(2), 193–205 (2001). <https://doi.org/10.1287/moor.26.2.193.10561>
3. Charina, M., Conti, C., Sauer, T.: Regularity of multivariate vector subdivision schemes. *Numer. Algorithms* **39**, 97–113 (2005). <https://doi.org/10.1007/s11075-004-3623-z>
4. Donoho, D.L., Tanner, J.: Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.* **22**, 1–53 (2009). <https://doi.org/10.1090/S0894-0347-08-00600-0>
5. Ferrez, J.-A., Fukuda, K., Liebling, T.M.: Solving the fixed rank convex quadratic maximization in binary variables by a parallel zonotope construction algorithm. *Eur. J. Oper. Res.* **166**(1), 35–50 (2005). <https://doi.org/10.1016/j.ejor.2003.04.011>
6. Gielen, R., Lazar, M.: On stability analysis methods for large-scale discrete-time systems. *Autom. J. IFAC* **55**, 6–72 (2015). <https://doi.org/10.1016/j.automatca.2015.02.034>
7. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM* **42**(6), 1115–1145 (1995). <https://doi.org/10.1145/227683.227684>
8. Gorskaya, E.S.: Approximation of convex functions by projections of polyhedra. *Moscow Univ. Math. Bull.* **65**(5), 196–203 (2010). <https://doi.org/10.3103/S0027132210050049>
9. Guglielmi, N., Protasov, V.Y.: Exact computation of joint spectral characteristics of linear operators. *Found. Comput. Math.* **13**(1), 37–97 (2013). <https://doi.org/10.1007/s10208-012-9121-0>
10. Guglielmi, N., Protasov, V.Y.: Invariant polytopes of sets of matrices with applications to regularity of wavelets and subdivisions. *SIAM J. Matrix Anal. Appl.* **37**(1), 18–52 (2016). <https://doi.org/10.1137/15M1006945>

11. Guglielmi, N., Wirth, F., Zennaro, M.: Complex polytope extremality results for families of matrices. *SIAM J. Matrix Anal. Appl.* **27**, 721–743 (2005). <https://doi.org/10.1137/040606818>
12. Guglielmi, N., Zennaro, M.: Balanced complex polytopes and related vector and matrix norms. *J. Convex Anal.* **14**, 729–766 (2007)
13. Guglielmi, N., Zennaro, M.: Finding extremal complex polytope norms for families of real matrices. *SIAM J. Matrix Anal. Appl.* **31**(2), 602–620 (2009). <https://doi.org/10.1137/080715718>
14. Guglielmi, N., Zennaro, M.: Canonical construction of polytope Barabanov norms and antinorms for sets of matrices. *SIAM J. Matrix Anal. Appl.* **36**(2), 634–655 (2015). <https://doi.org/10.1137/140962814>
15. Gurvits, L.: Stability of discrete linear inclusions. *Linear Algebra Appl.* **231**, 47–85 (1995). [https://doi.org/10.1016/0024-3795\(95\)90006-3](https://doi.org/10.1016/0024-3795(95)90006-3)
16. Hästad, J.: Clique is hard to approximate within $V^{1-\epsilon}$. *Acta Math.* **182**, 105–142 (1999). <https://doi.org/10.1007/BF02392825>
17. Hästad, J.: Some optimal inapproximability results. *J. ACM* **48**, 798–859 (2001). <https://doi.org/10.1145/502090.502098>
18. Hueter, I.: Limit theorems for the convex hull of random points in higher dimensions. *Trans. Amer. Math. Soc.* **351**(11), 4337–4363 (1999). <https://doi.org/10.1090/S0002-9947-99-02499-X>
19. Jungers, R.: The joint spectral radius. Theory and applications, Lecture Notes in Control and Information Sciences. Springer, ISBN: 978-3-540-95980-9 (2009)
20. Kozyakin, V.S.: Structure of extremal trajectories of discrete linear systems and the finiteness conjecture. *Autom. Remote Control* **68**, 174–209 (2007). <https://doi.org/10.1134/S0005117906040171>
21. Kozyakin, V.S.: Iterative building of Barabanov norms and computation of the joint spectral radius for matrix sets. *Discrete Contin. Dyn. Syst. Ser. B* **14**, 143–158 (2010)
22. Maesumi, M.: Optimal norms and the computation of joint spectral radius of matrices. *Linear Algebra Appl.* **428**(10), 2324–2338 (2008)
23. Mejstrik, T.: Improved invariant polytope algorithm and applications. *ACM Trans. Math. Softw.* **46**(3), 1–26 (2020). <https://doi.org/10.1145/3408891>
24. Mejstrik, T.: Matlab toolbox for work with subdivision schemes and joint spectral radius, <https://gitlab.com/tommsch/ttoolboxes>
25. Mejstrik, T., Hollomey, C.: TTEST framework—unit test framework for Matlab/Octave, <https://gitlab.com/tommsch/TTEST>
26. Plischke, E., Wirth, F.: Duality results for the joint spectral radius and transient behaviour. *Linear Algebra Appl.* **428**, 2368–2384 (2008). <https://doi.org/10.1109/CDC.2005.1582512>
27. Protasov, V.Yu.: The joint spectral radius and invariant sets of linear operators. *Fund. i prikladnaya Mat.* **2**(1), 205–231 (1996). <http://mi.mathnet.ru/eng/fpm/v2/i1/p205>
28. Protasov, V.Y.: The Euler binary partition function and subdivision schemes. *Math. Comput.* **86**, 1499–1524 (2017). <https://doi.org/10.1090/mcom/3128>
29. Protasov, V.Y.: The Barabanov norm is generically unique, simple, and easily computed. *SIAM J. Optim. Control* **60**(4), 2246–2267 (2022). [arXiv: 2109.12159](https://arxiv.org/abs/2109.12159)
30. Sylvester, J.J.: Problem 1491, *The Educational Times* (London), April, 1–28 (1864)
31. Wirth, F.: The generalized spectral radius and extremal norms. *Linear Algebra Appl.* **342**, 17–40 (2002). [https://doi.org/10.1016/S0024-3795\(01\)00446-3](https://doi.org/10.1016/S0024-3795(01)00446-3)
32. Wu, Z., He, Q.: Optimal switching sequence for switched linear systems. *SIAM J. Control Optim.* **58**(2), 1183–1206 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.