Calculating the EHZ Capacity of Polytopes

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Stefan Krupp

aus Linz/Rhein

Köln, 2020

Berichterstatter (Gutachter):

Prof. Dr. Frank Vallentin,	Universität zu Köln
Prof. Hansjörg Geiges, Ph.D. (Cantab)	Universität zu Köln

Tag der mündlichen Prüfung: 18.02.2021

Preface

There is a widely held prejudice that mathematicians are strange people who work in solitude and reduce their social activities to a minimum. Ironically, the opposite is usually the case. Mathematics thrives on the exchange of ideas and cooperations. This thesis, for example, would not have been possible without the advice, help and support from various people. Therefore, I would like to take the opportunity to properly express my gratitude.

First and foremost, I thank my supervisor Frank Vallentin for so many reasons. He guided me not only through almost four years of research but also through the writing of my Bachelor and Master thesis. In the process, he was always open for meetings and discussions during which he gave me lots of advice and exciting ideas. Furthermore, it was he who introduced me to convex optimization through his lectures and thus made this thesis possible in the first place. I appreciate all you have done for me.

Next, I would like to thank my collaborator and co-author Daniel Rudolf. I enjoyed our joint work and in particular all the productive meetings in which we assigned "homework" to each other. In particular, thank you for your explanations about billiards and your results, which are essential for a large part of this thesis.

I also want to thank the principal investigators Alberto Abbondandolo, Peter Albers and Frank Vallentin for giving me the opportunity to participate in this exciting project and for the helpful advices that improved this thesis. Additionally, I would like to express my gratitude towards the Deutsche Forschungsgemeinschaft for funding this project.

A big thanks goes out to my colleagues Davi de Castro Silva, Maria Dostert, Anna Gundert, Arne Heimendahl, Frederic von Heymann, Annette Koenen, Jan Rolfes and Marc Zimmermann for creating such a wonderful working environment that made me enjoy my time in the office. In particular, I thank Maria and Jan for the warm welcome to our shared office room and for the collaboration on the Oberwolfach snapshot. I had so much fun tipping the scales between you. Moreover, I thank Davi for the gorgeous anteater figurine that cheered me up so many times during the last years.

Also, I would like to thank Simon Brebeck, who implemented the graphical user interface for our Minkowski billiard algorithm during his internship. I wish you best of luck for your future career.

Furthermore, I thank Hansjörg Geiges, Peter Littelmann, Frank Vallentin and Marc Zimmermann for agreeing to form the examination committee and for assessing this thesis.

I appreciate that so many friends and colleagues agreed to proofread parts of this thesis. More precisely, I thank Davi de Castro Silva, Maria Dostert, Greta Fischer, Anna Gundert, Arne Heimendahl, Frederic von Heymann, Christina Röhrig, Jan Rolfes and Marc

Zimmermann.

A special thanks goes out to my family, most notably my parents Gisela and Jürgen, and to my friends for their support and their encouragement. I owe you a lot. Finally, I thank my loving girlfriend Janine for all the small and big things she is doing for me. I cannot emphasize enough how happy I am to have you in my life.

Kurzfassung

In dieser Arbeit betrachten wir die symplektische Geometrie. Dieser eher junge Forschungsbereich stößt seit dem späten 20. Jahrhundert auf zunehmend Interesse aufgrund einiger namhafter Resultate wie dem berühmten Non-squeezing Theorem von Gromov. Eine Besonderheit dieses Theorems ist die Tatsache, dass seine Aussage genau dann wahr ist, wenn es globale symplektische Invarianten gibt, die bestimmte Eigenschaften erfüllen. Diese globalen Invarianten werden symplektische Kapazitäten genannt und ihre Existenz ist eine nicht triviale Tatsache. Mittlerweile sind mehrere Konstruktionen bekannt, jedoch wurde die Berechnung symplektischer Kapazitäten bislang kaum betrachtet. In dieser Arbeit beschäftigen wir uns mit diesem Gesichtspunkt für eine bestimmte symplektische Kapazität, nämlich der Ekeland-Hofer-Zehnder Kapazität, welche jedem konvexen Körper in $\mathbb{R}^{2n}$ eine nichtnegative Zahl oder ∞ zuordnet.

Um die Ekeland-Hofer-Zehnder Kapazität von Mengen der Form $K \times T$ zu berechnen, wobei K und T konvexe Mengen sind, verwenden wir einen Ansatz, der auf Minkowski Billards basiert. Genauer bauen wir auf einem Algorithmus von Alkoumi und Schlenk auf, der zweidimensionale konvexe Mengen K betrachtet und annimmt, dass T die euklidische Einheitskugel ist. Einerseits bearbeiten wir diesen Algorithmus so, dass auch konvexe Mengen K mit beliebiger Dimension betrachtet werden können. Andererseits verallgemeinern wir den Ansatz von Alkoumi und Schlenk vom euklidischen Fall (wo T die euklidische Einheitskugel ist) zum Minkowski Fall (wo T eine beliebige konvexe Menge ist). Insbesondere betrachten wir die Situation, in der K und T Polytope sind.

Abgesehen von diesem Ansatz betrachten wir ein Resultat von Abbondandolo und Majer, welches die Ekeland-Hofer-Zehnder Kapazität als ein Maximierungsproblem formuliert. Dieses Maximierungsproblem ist ein Hybrid aus einem quadratischen Zuordnungsproblem und einem quadratischen Programm mit nichtkonvexer Zielfunktion. Wir verwenden verschiedene Optimierungstechniken, um obere und untere Schranken an die Ekeland-Hofer-Zehnder Kapazität von Polytopen zu finden. Unter Anderem erhalten wir eine sehr gute obere Schranke, die für viele kleinere Probleminstanzen dem Optimalwert entspricht, jedoch eine schnell wachsende Laufzeit aufweist.

Abstract

In this thesis we are concerned with symplectic geometry. This rather new field of research gained much interest due to some famous results in the late 20th century, such as the celebrated non-squeezing theorem by Gromov. An intriguing feature of this theorem is, that it is true if and only if there are global symplectic invariants that satisfy certain properties. These global invariants are called symplectic capacities and their existence is a nontrivial fact. By now, several constructions are known but the computation of symplectic capacities has not received much attention. In this thesis, we address this aspect for a certain symplectic capacity, namely the Ekeland-Hofer-Zehnder capacity, which maps a nonnegative number or infinity to every convex body in $\mathbb{R}^{2n}$.

To compute the Ekeland-Hofer-Zehnder capacity of sets of the form $K \times T$, where both K and T are convex sets, we use an approach based on Minkowski Billiards. More precisely, we build on an algorithm by Alkoumi and Schlenk, that is formulated for the case where K is a two-dimensional convex set and T is the Euclidean unit ball. On the one hand, we adapt this algorithm to allow convex sets K with arbitrary dimension. On the other hand, we generalize the approach by Alkoumi and Schlenk from the Euclidean setting (where T is the Euclidean unit ball) to the Minkowski setting (where T is an arbitrary convex set). In particular, we consider the case where both K and T are polytopes.

Aside from this approach we consider a formulation of the Ekeland-Hofer-Zehnder capacity as a maximization problem, which is due to Abbondandolo and Majer. This maximization problem is a hybrid of a quadratic assignment problem and a quadratic program with non-convex objective function. We employ different optimization techniques and obtain upper and lower bounds on the Ekeland-Hofer-Zehnder capacity of polytopes. Amongst others, we obtain a very good upper bound that is equal to the exact value for many small problem instances at the cost of a rapidly increasing running time.

Contents

Chapter One

Introduction

What is symplectic geometry and why are people interested in it? To answer this question we take a look into the past. Even though symplectic geometry is a rather new subject, its roots are old and date back to the early 19th century. In 1808 Joseph-Louis Lagrange used classical mechanics to study the behaviour of planets. While doing so, he came up with a simple way to formulate classical mechanics . Later William Rowan Hamilton built up on Lagrange's formulation and established that if $q(t) \in \mathbb{R}^3$ denotes the position and $p(t) \in \mathbb{R}^3$ the momentum of a planet at the time t, then the motion of this planet is given by the differential equations

$$\frac{dp}{dt} = \frac{\partial H}{\partial q}, \ \frac{dq}{dt} = -\frac{\partial H}{\partial p},$$

where $H(q(t), p(t), t)$ is the energy of the planet at the time t [100]. We can think of symplectic geometry as a mathematical formalization of Hamilton's formulation. More precisely, symplectic geometry is about manifolds that are equipped with a skew-symmetric, nondegenerate bilinear form, which we call a symplectic form and denote by ω. The prime example of a symplectic form is the standard symplectic form

$$\omega_0(u, v) = u^T \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} v \text{ for } u, v \in \mathbb{R}^{2n}.$$

While symplectic geometry did not receive much attention for a long time, it is now an active and thriving subject with connections to many other areas, for instance classical and quantum mechanics, optics, algebraic geometry and low-dimensional topology [12], [30], [40], [44], [74], [95]. Multiple celebrated discoveries in the last few decades are the reason for this renewed interest. One of these results is Gromov's non-squeezing theorem [42].

Theorem 1.0.1 (Non-squeezing Theorem)**.** *Let $B_r(0) \subseteq \mathbb{R}^{2n}$ be the ball of radius r and*

$$Z_R(0) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \colon x_1^2 + y_1^2 \leqslant R^2\},$$

with $r, R \geqslant 0$. There is an injective symplectic map $\varphi \colon B_r(0) \to Z_R(0)$ if and only if $r \leqslant R$.

The term "symplectic map" refers to a differentiable map which preserves the structure given by the corresponding symplectic form.

One important aspect of Gromov's theorem is its relation to certain symplectic invariants which are called symplectic capacities. More precisely, a *symplectic capacity* c maps each subset of $\mathbb{R}^{2n}$ to a nonnegative number or to infinity such that the following properties hold for every subset $A \subseteq \mathbb{R}^{2n}$:

- $c(A) \leqslant c(B)$ for every set B with $A \subseteq B \subseteq \mathbb{R}^{2n}$,
- $c(\varphi(A)) = c(A)$ for every symplectic diffeomorphism φ,
- $c(\alpha A) = \alpha^2 c(A)$ for every $\alpha \in \mathbb{R}$, $\alpha > 0$,
- $c(B_r(0)) = c(Z_r(0)) = \pi r^2$,

where $B_r(0)$ and $Z_r(0)$ are defined as in Theorem 1.0.1.

The existence of a symplectic capacity is sufficient to prove Gromov's non-squeezing theorem, but whether there is a symplectic capacity is not obvious. Nevertheless, several constructions have been found [42], [49], [50]. On the other hand, the question of how to compute $c(A)$ for a given symplectic capacity c and $A \subseteq \mathbb{R}^{2n}$ did not receive much attention. So, exact values are only known for a few special cases [81].

Furthermore, the study of invariances in symplectic geometry is of particular meaning. In 1882, Darboux showed that any two symplectic manifolds with the same dimension are locally isomorphic [29]. This rules out any local symplectic invariants. Thus, the question arises of what global invariants look like.

In Chapter 2 we revisit the content of this introductory section to give a more rigorous, albeit basic, introduction to symplectic geometry. Most notably, this includes the definition of the Ekeland-Hofer-Zehnder capacity. This particular symplectic capacity is the central object of this thesis. Furthermore, we take a look at Hamiltonian mechanics and a famous conjecture by Mahler and see how these topics relate to the symplectic setting.

Even though the investigation of the Ekeland-Hofer-Zehnder capacity is a problem that stems from symplectic geometry, we mostly make use of optimization techniques in this thesis. We dedicate Chapter 3 to the corresponding preliminaries. Aside of an introduction to conic optimization and some prominent special cases, we take a look at a particular optimization problem, namely the quadratic assignment problem.

In Chapter 4 we consider convex sets of the form $K \times T$, where $K, T \subseteq \mathbb{R}^n$ are convex as well. The Ekeland-Hofer-Zehnder capacity of $K \times T$ is closely related to the concept of closed billiards. More precisely, K plays the role of the billiard table and the set T determines the bouncing rule. First, we study some theoretical aspects of this approach. Then, we consider the Euclidean setting, i.e. where K is a polytope and T is the unit ball. We provide a polynomial-time algorithm to compute an upper bound on the Ekeland-Hofer-Zehnder capacity in this setting. The input of this algorithm consists of the vertices and the facets of K. Afterwards, we examine the Minkowski setting, i.e. where both K and T are polytopes. In this case we give a polynomial-time algorithm to compute the Ekeland-Hofer-Zehnder capacity if both K and T are two-dimensional. The input consists of the vertices and the facets of K and T. This chapter is based on [65] and [66], which is currently

in preparation. Both of these publications are joint work with Daniel Rudolf from Ruhr-University Bochum.

In Chapter 5 we step away from convex sets of the form $K \times T$ and consider arbitrary $2n$-dimensional polytopes instead. We use a result by Abbondandolo and Majer that formulates the Ekeland-Hofer-Zehnder capacity as an optimization problem [1]. Furthermore, we discuss the challenges that arise with this formulation and apply different strategies to compute upper and lower bounds. In particular, a strategy that relies on semidefinite relaxation yields remarkable upper bounds. Therefore, we elaborate on the quality of our results by employing a rank minimization technique. In some cases this approach yields a proof that our upper bound is tight.

Chapter Two

Basics of symplectic geometry

In this chapter we give the basic notions and concepts of symplectic geometry. We do so by mainly following [51]. Usually literature on symplectic geometry discusses symplectic manifolds. In this thesis we concentrate on symplectic vector spaces instead, since the manifold we are considering is $\mathbb{R}^{2n}$.

The goal of this chapter is to define the central object of this thesis, namely the Ekeland-Hofer-Zehnder capacity, and to provide some motivation. The techniques used in later chapters are not of symplectic nature and therefore do not appear in Chapter 2.

2.1 Symplectic vector spaces

We start our introduction to symplectic geometry by defining our setup. Let V be an n-dimensional real vector space and $\omega\colon V \times V \to \mathbb{R}$ a bilinear form that is

- skew-symmetric: $\omega(u,v) = -\omega(v,u)$ for all $u, v \in V$ and
- nondegenerate: for every $u \in V\backslash\{0\}$ there is $v \in V$ such that $\omega(u,v) \neq 0$.

We call ω a *symplectic form* and the pair (V,ω) a *symplectic vector space*. The simplest example of a symplectic vector space is $(\mathbb{R}^{2n}, \omega_0^{2n})$, where for every $u, v \in \mathbb{R}^{2n}$ we define

$$\omega_0^{2n}(u,v) = u^T J_{2n} v \quad \text{with} \quad J_{2n} = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}.$$

Here, I_n denotes the $(n \times n)$-identity matrix. We write ω_0 and J instead of ω_0^{2n} and J_{2n} if the dimension is clear from the context. On the one hand, we observe that J is regular because $\det J = 1$. Thus, ω_0 is nondegenerate. On the other hand, $J^T = -J$ which implies skew-symmetry:

$$\omega_0(u,v) = u^T J v = (u^T J v)^T = v^T J^T u = -\omega_0(v,u) \;\; \forall u, v \in V.$$

The pair $(\mathbb{R}^{2n}, \omega_0)$ is called the *standard symplectic vector space* and in fact it is more than just a simple example. Every symplectic vector space looks like $(\mathbb{R}^{2n}, \omega_0)$ provided we use the correct basis.

Proposition 2.1.1. *[51] Let (V, ω) be a finite-dimensional symplectic vector space. Then it has even dimension $2n$ and there is a basis $e_1, \ldots, e_n, f_1, \ldots, f_n$ of V such that for every $i, j \in \{1, \ldots, n\}$:*

- $\omega(e_i, e_j) = 0$,
- $\omega(f_i, f_j) = 0$,
- $\omega(f_i, e_j) = \delta_{ij}$.

Proof. If $V = \{0\}$ there is nothing to prove. So, we assume there is $e_1 \in V \backslash \{0\}$. Since ω is nondegenerate, we can find $w \in V$ such that $\omega(e_1, w) \neq 0$. Letting $f_1 = w/\omega(e_1, w)$, we get $\omega(e_1, f_1) = 1$. Note that f_1 is not a multiple of e_1 because skew-symmetry and bilinearity of ω imply for every $v \in V$ and $\alpha \in \mathbb{R}$

$$\omega(v, \alpha v) = -\omega(\alpha v, v) = -\omega(v, \alpha v)$$

and hence $\omega(v, \alpha v) = 0$. In other words, e_1 and f_1 are linearly independent. If $\dim V = 2$, the proof is finished. Otherwise, we consider the subspaces

$$\begin{aligned} E &= \operatorname{span}\{e_1, f_1\}, \\ E^{\perp} &= \{u \in V \colon \omega(u, v) = 0 \ \ \forall\, v \in E\}. \end{aligned}$$

Next we would like to argue that

$$\dim E + \dim E^{\perp} = \dim V. \tag{2.1}$$

To this end, we let $\{b_1, \ldots, b_r\}$ be a basis of $E^{\perp}$, $r = \dim E^{\perp}$. We can extend this to a basis $\{b_1, \ldots, b_s\}$ of V, $s = \dim V$. Furthermore, we define a matrix $M \in \mathbb{R}^{2 \times s}$ by

$$\left.\begin{aligned} M_{1,j} &= \omega(e_1, b_j) \\ M_{2,j} &= \omega(f_1, b_j) \end{aligned}\right\} \text{ for all } j \in \{1, \ldots, s\}.$$

On the one hand, a simple calculation yields that a vector $y \in \mathbb{R}^s$ is in the kernel of M if and only if $\sum_j y_j b_j$ is contained in $E^{\perp}$. Because $\{b_1, \ldots, b_r\}$ is a basis of $E^{\perp}$, we conclude that $y_{r+1} = \ldots = y_s = 0$ and that $\dim \ker M = r$.

On the other hand, we observe that the rows of M are linearly independent. To see this, we take $\lambda_1, \lambda_2 \in \mathbb{R}$ such that

$$0 = \lambda_1 \begin{pmatrix} \omega(e_1, b_1) \\ \vdots \\ \omega(e_1, b_s) \end{pmatrix} + \lambda_2 \begin{pmatrix} \omega(f_1, b_1) \\ \vdots \\ \omega(f_1, b_s) \end{pmatrix} = \begin{pmatrix} \omega\left(\lambda_1 e_1 + \lambda_2 f_1, b_1\right) \\ \vdots \\ \omega\left(\lambda_1 e_1 + \lambda_2 f_1, b_s\right) \end{pmatrix}.$$

Since $\{b_1, \ldots, b_s\}$ is a basis of V, we find that

$$\omega(\lambda_1 e_1 + \lambda_2 f_1, v) = 0$$

holds for every $v \in V$. Now nondegeneracy implies that $\lambda_1 e_1 + \lambda_2 f_1 = 0$ and we get $\lambda_1 = \lambda_2 = 0$ because $\{e_1, f_1\}$ is a basis of E. Thus, $\operatorname{rk} M = 2$. Identity (2.1) follows by the rank-nullity theorem if we plug in $r = \dim E^{\perp}$, $s = \dim V$ and $2 = \dim E$.

Because of $\omega(e_1, f_1) = 1$, we have $E \cap E^\perp = \{0\}$. Together with (2.1) this means that the subspaces $E, E^\perp$ decompose V, i.e. $V = E \oplus E^\perp$. We can apply the same reasoning as at the start of the proof to $E^\perp$ inductively and conclude the proof. □

Now we have an acurate idea of what symplectic spaces look like. Thus, we usually focus on the standard symplectic space $(\mathbb{R}^{2n}, \omega_0)$ throughout this thesis.

2.2 Symplectic maps and the action functional

Next, we turn our attention to maps on symplectic vector spaces. Since a symplectic form provides a symplectic vector space with some structure, it is instructive to consider maps that leave this structure invariant. This motivates the following definition.

Definition 2.2.1. *For $i \in \{1, 2\}$ let $U_i \subseteq \mathbb{R}^{2n}$ be an open set and let $\omega_i \colon \mathbb{R}^{2n} \times \mathbb{R}^{2n} \to \mathbb{R}$ be a symplectic form. Furthermore, let $\varphi \colon U_1 \to U_2$ be a differentiable map that for every $x \in U_1$ has the property*

$$\omega_1(u, v) = \omega_2(D_\varphi(x)u, D_\varphi(x)v) \quad \forall u, v \in \mathbb{R}^{2n}, \tag{2.2}$$

where $D_\varphi(x)$ is the Jacobian of φ at x. Then we say that φ is symplectic.

It is an immediate consequence of Proposition 2.1.1 that between every two symplectic vector spaces with same dimension there is a linear symplectic map (more precisely a symplectic isomorphism). This symplectic map is a change of basis according to Proposition 2.1.1.

A distinctive property of a symplectic diffeomorphism $\varphi \colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is that it preserves the volume. More precisely, we consider the standard symplectic vector space $(\mathbb{R}^{2n}, \omega_0)$. Identity (2.2) then reads for every $x \in \mathbb{R}^{2n}$:

$$u^T J v = u^T D_\varphi(x)^T J D_\varphi(x) v \quad \forall u, v \in \mathbb{R}^{2n}. \tag{2.3}$$

This is equivalent to

$$D_\varphi(x)^T J D_\varphi(x) = J$$

and immediately implies $\det D_\varphi(x) \in \{-1, +1\}$. In fact, $\det D_\varphi(x) = 1$ holds (see for instance [33]). Thus, $\operatorname{vol} \varphi(U) = \operatorname{vol} U$ for every $U \subseteq \mathbb{R}^{2n}$.

In addition to the symplectic form ω_0 and the volume there is another functional that is left invariant under symplectic diffeomorphisms. It is called the action functional. To state it properly, we use the language of differential forms. The concepts and notions that we utilize are common and can be found, for instance, in [24], [39] and [56].

We let $x_i, y_i \colon \mathbb{R}^{2n} \to \mathbb{R}$ for $i \in \{1, \ldots, n\}$ be coordinate functions defined by

$$x_i(u) = u_i, \; y_i(u) = u_{n+i}$$

for $u \in \mathbb{R}^{2n}$. We can express the standard symplectic form as

$$\omega_0 = \sum_{i=1}^{n} dy_i \wedge dx_i.$$

We can verify this expression using the definition of the wedge product. For $u, v \in \mathbb{R}^{2n}$ we have

$$\left(\sum_{i=1}^{n} dy_i \wedge dx_i\right)(u,v) = \sum_{i=1}^{n} (dy_i \wedge dx_i)(u,v) = \sum_{i=1}^{n} dy_i(u)dx_i(v) - dy_i(v)dx_i(u)$$
$$= \sum_{i=1}^{n} u_{n+i}v_i - u_i v_{n+i} = u^T J v.$$

There is another common way to express the standard symplectic form, namely $\omega_0 = d\lambda$, where

$$\lambda = \sum_{i=1}^{n} y_i dx_i$$

is the Liouville 1-form. This is due to the following calculation:

$$d\lambda = d\left(\sum_{i=1}^{n} y_i dx_i\right) = \sum_{i=1}^{n} d(y_i dx_i) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} \frac{\partial y_i}{\partial z_j} dx_j \wedge dx_i + \sum_{j=1}^{n} \frac{\partial y_i}{\partial z_{n+j}} dy_j \wedge dx_i\right)$$
$$= \sum_{i=1}^{n} dy_i \wedge dx_i = \omega_0.$$

We are now ready to state the definition of the *action functional*.

Definition 2.2.2. *Let $\gamma \subseteq \mathbb{R}^{2n}$ be a closed curve. Then*

$$\mathbb{A}(\gamma) = \int_{\gamma} \lambda$$

is called the action of γ.

As mentioned before, applying symplectic diffeomorphisms leaves the action functional invariant. This is easy to show if we utilize differential forms. Therefore, we state the proof for the sake of completeness.

Proposition 2.2.3. *[51] Let $\gamma \subseteq \mathbb{R}^{2n}$ be a closed curve and $\varphi\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be a symplectic diffeomorphism. Then*

$$\mathbb{A}(\gamma) = \mathbb{A}(\varphi(\gamma)).$$

Proof. First, we reformulate (2.3) using the pullback φ^* to get

$$\omega_0 = \varphi^* \omega_0.$$

Indeed, this equation is equivalent to (2.3):

$$\varphi^*\omega_0 = \sum_{i=1}^{n}\left(\sum_{j=1}^{2n} \frac{\partial \varphi_{n+i}}{\partial z_j} dz_j\right) \wedge \left(\sum_{j=1}^{2n} \frac{\partial \varphi_i}{\partial z_j} dz_j\right) = \sum_{i=1}^{n} \sum_{1 \leqslant j,\ell \leqslant 2n} \frac{\partial \varphi_{n+i}}{\partial z_j} \frac{\partial \varphi_i}{\partial z_\ell} dz_j \wedge dz_\ell.$$

Plugging in some points $u, v, x \in \mathbb{R}^{2n}$ we get

$$\begin{aligned}(\varphi^*\omega_0)_x(u,v) = \sum_{i=1}^{n}\sum_{1\leqslant j,\ell\leqslant 2n} D_\varphi(x)_{n+i,j}D_\varphi(x)_{i,\ell}(u_j v_\ell - u_\ell v_j) &= u^T D_\varphi(x)^T J D_\varphi(x) v \\ &= \omega_0(D_\varphi(x)u, D_\varphi(x)v).\end{aligned}$$

The second equality in this equation is a simple but lengthy calculation. Therefore, we omit the details.

Next we consider the 1-form $\lambda - \varphi^*\lambda$. Because the exterior derivative d and the pullback φ^* commute, we have

$$d(\lambda - \varphi^*\lambda) = d\lambda - \varphi^* d\lambda = \omega_0 - \varphi^*\omega_0 = \omega_0 - \omega_0 = 0.$$

Thus, the 1-form $\lambda - \varphi^*\lambda$ is exact and according to the Poincaré lemma [69] there is a function $F\colon \mathbb{R}^{2n} \to \mathbb{R}$ such that

$$\lambda - \varphi^*\lambda = dF.$$

Now Stokes' theorem [69] implies

$$\int_\gamma \lambda - \varphi^*\lambda = \int_\gamma dF = \int_{\partial\gamma} F = 0,$$

where the last equality holds because γ is a closed curve. This completes the proof as

$$\mathbb{A}(\gamma) = \int_\gamma \lambda = \int_\gamma \varphi^*\lambda = \int_{\varphi(\gamma)} \lambda = \mathbb{A}(\varphi(\gamma)).$$

□

In the upcoming chapters we consider problems in which the action functional takes a significant role. Therefore, it is beneficial to examine it closer. Above, we state the action functional using differential forms. Next, we want to reformulate this definition such that we receive an expression for $\mathbb{A}(\gamma)$ that does not rely on the notion of differential forms. The trick is to parametrize the closed curve γ by a function $f\colon [0,a] \to \mathbb{R}^{2n}$ with $f(0) = f(a)$ for some $a \in \mathbb{R}_{\geqslant 0}$. This idea is mentioned in [51]. Here, we carry out the calculation in more detail.

Proposition 2.2.4. *Let $f\colon [0,a] \to \mathbb{R}^{2n}$ be a differentiable function with $f(0) = f(a)$ for some $a \in \mathbb{R}_{\geqslant 0}$. Then*

$$\mathbb{A}(f) := \mathbb{A}(f([0,a])) = -\frac{1}{2}\int_0^a \dot{f}(t)^T J f(t) dt.$$

Proof. Let $f_i = x_i \circ f$ and $f_{n+i} = y_i \circ f$ for $i \in \{1,\dots,2n\}$. In other words $f_i(t)$ is the ith coordinate of $f(t)$ for $i \in \{1,\dots,2n\}$ and $t \in [0,a]$. We perform a change of variables which means

$$\mathbb{A}(f) = \int_{f([0,a])} \lambda = \int_0^a f^*\lambda = \int_0^a \sum_{i=1}^{n} f_{n+i}(t) f_i'(t) dt. \tag{2.4}$$

Next, we switch summation with integration and integrate by parts.

$$\mathbb{A}(f) = \sum_{i=1}^{n} \int_0^a f_{n+i}(t) f_i'(t) dt = \sum_{i=1}^{n} \left[f_{n+i}(a) f_i(a) - f_{n+i}(0) f_i(0) - \int_0^a f_{n+i}'(t) f_i(t) dt \right] \tag{2.5}$$

$$= -\sum_{i=1}^{n} \int_0^a f_{n+i}'(t) f_i(t) dt.$$

The last equation holds since $f(0) = f(a)$. Adding (2.4) and (2.5), we find

$$2\mathbb{A}(f) = \int_0^a \sum_{i=1}^{n} f_{n+i}(t) f_i'(t) - f_{n+i}'(t) f_i(t) dt = -\int_0^a \dot{f}(t)^T J f(t) dt. \qquad \square$$

In the statement of Proposition 2.2.4 we obviously require that the parametrization f is differentiable since the differential of f appears in the statement as well. However, we can consider piecewise differentiable curves instead. Say we have points $0 = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_m = a$ and a continuous function f that is differentiable on (t_{j-1}, t_j) for every $j \in \{1, \ldots, m\}$. Again, we require $f(0) = f(a)$. Then we can extend Proposition 2.2.4 to get

$$\mathbb{A}(f) = -\sum_{j=1}^{m} \frac{1}{2} \int_{t_{j-1}}^{t_j} \dot{f}(t)^T J f(t) dt.$$

The proof is similar with the exception that we need to split up the integral. Then (2.4) becomes

$$\mathbb{A}(f) = \sum_{j=1}^{m} \int_{f([t_{j-1}, t_j])} \lambda = \sum_{j=1}^{m} \int_{t_{j-1}}^{t_j} \sum_{i=1}^{n} f_{n+i}(t) f_i'(t) dt$$

and instead of (2.5) we have

$$\mathbb{A}(f) = \sum_{j=1}^{m} \sum_{i=1}^{n} \left[f_{n+i}(t_j) f_i(t_j) - f_{n+i}(t_{j-1}) f_i(t_{j-1}) - \int_{t_{j-1}}^{t_j} f_{n+i}'(t) f_i(t) dt \right],$$

where the term

$$\sum_{j=1}^{m} \sum_{i=1}^{n} f_{n+i}(t_j) f_i(t_j) - f_{n+i}(t_{j-1}) f_i(t_{j-1})$$

is a telescopic sum and vanishes like in the proof of Proposition 2.2.4 due to $f(0) = f(a)$.

2.3 Symplectic capacities

We now introduce the object that we primarily study in this thesis, namely symplectic capacities. Usually, a symplectic capacity is defined as a map c that associates to each symplectic manifold (i.e. a manifold equipped with a symplectic form) a nonnegative number or infinity such that certain conditions hold (see Definition 2.3.1). As mentioned earlier, we are interested in a more specific setting. Therefore, we state this definition for the case where all appearing manifolds are open subsets of $\mathbb{R}^{2n}$.

Definition 2.3.1. *Let $\mathcal{M}_{2n}$ be the set of all open subsets of $\mathbb{R}^{2n}$ and let Ω_{2n} be the set of all symplectic forms on $\mathbb{R}^{2n}$. A map*

$$c\colon \mathcal{M}_{2n} \times \Omega_{2n} \longrightarrow \mathbb{R}_{\geqslant 0} \cup \{\infty\}$$

is called a symplectic capacity if it has the following properties:

(i) If $A \in \mathcal{M}_{2n}$ is equipped with $\omega \in \Omega_{2n}$ and $B \in \mathcal{M}_{2n}$ is equipped with $\tau \in \Omega_{2n}$ such that there is an injective symplectic map $\varphi\colon A \to B$, then:

$$c(A,\omega) \leqslant c(B,\tau).$$

(ii) If $A \in \mathcal{M}_{2n}$, $\omega \in \Omega_{2n}$ and $\alpha \in \mathbb{R}\backslash\{0\}$, then:

$$c(A,\alpha\omega) = |\alpha| c(A,\omega).$$

(iii) Let $B_1(0)$ be the Euclidean unit ball and let

$$Z_1(0) = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n\colon x_1^2 + y_1^2 \leqslant 1\}$$

be an infinitely long cylinder. Then:

$$c(\operatorname{int} B_1(0), \omega_0) = \pi = c(\operatorname{int} Z_1(0), \omega_0).$$

An immediate implication of this definition is that $c(A,\omega) = c(B,\tau)$ if there is a symplectic diffeomorphism $\varphi\colon A \to B$. One simply applies the monotonicity property (i) to the symplectic maps φ and φ^{-1}. Furthermore, we observe that $c(A,\omega) \leqslant c(B,\omega)$ for every $\omega \in \Omega_{2n}$ whenever $A \subseteq B$. To see this, we apply the monotonicity property to the map $\varphi = \mathrm{id}$. This map is injective, differentiable and its Jacobian is the identity matrix. Thus, φ is symplectic. A third property that we deduce from Definition 2.3.1 is the 2-homogeneity.

Lemma 2.3.2. *[51] Let $A \subseteq \mathbb{R}^{2n}$ be open and let ω be a symplectic form on $\mathbb{R}^{2n}$. Then for every $\alpha \in \mathbb{R}\backslash\{0\}$ and every symplectic capacity c we have*

$$c(\alpha A, \omega) = \alpha^2 c(A,\omega).$$

Proof. We define the map $\varphi\colon \alpha A \to A$ by

$$\varphi(x) = \frac{1}{\alpha} x \text{ for all } x \in \alpha A.$$

We see that φ is a diffeomorphism and its Jacobian is given by

$$D_\varphi(x) = \frac{1}{\alpha} I.$$

In particular, we have for every $x \in \alpha A$:

$$\omega(u, v) = \alpha^2 \omega(D_\varphi(x)u, D_\varphi(x)v) \ \ \forall\ u, v \in \mathbb{R}^{2n}.$$

So, φ is a symplectic diffeomorphism mapping αA to A, where we consider $(\alpha A, \omega)$ and $(A, \alpha^2\omega)$. This means that we can apply the monotonicity property with equality to get

$$c(\alpha A, \omega) = c(A, \alpha^2 \omega).$$

The statement now follows if we apply the conformality property (ii). □

With Lemma 2.3.2 we can extend the nontriviality property (iii). We have

$$c(\operatorname{int} B_r(0), \omega_0) = \pi r^2 = c(\operatorname{int} Z_r(0), \omega_0), \tag{2.6}$$

for every $r > 0$, where $B_r(0)$ is the Euclidean ball of radius r and center 0 and

$$Z_r(0) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \colon x_1^2 + y_1^2 \leqslant r^2\}.$$

We note that in dimension two the sets $Z_r(0)$ and $B_r(0)$ are equal for every radius $r > 0$. Therefore, the requirement (2.6) lets us suspect a relation between symplectic capacities and the two-dimensional volume. In fact, Siburg [92] proved that $c(D, \omega_0) = \operatorname{vol} D$ holds for every compact and connected domain $D \subseteq \mathbb{R}^2$ with smooth boundary. This however, is a special feature in dimension two. In every even dimension greater than two the set $Z_r(0)$ is unbounded for every $r > 0$. This means that the volume of $Z_r(0)$ is infinite which disqualifies the volume as a symplectic capacity.

The next property of symplectic capacities that we discuss is the continuity with respect to the Hausdorff metric. For each two sets $A, B \subseteq \mathbb{R}^{2n}$ we define the *Hausdorff metric* by

$$d_H(A, B) = \max\left\{ \sup_{a\in A} \inf_{b\in B} ||a - b||, \sup_{b\in B} \inf_{a\in A} ||a - b|| \right\}.$$

This continuity property is stated as an exercise in [74] and a proof can be found in [7]. Here, we give a slightly different proof.

Lemma 2.3.3. *Let $A \subseteq \mathbb{R}^{2n}$ be open, bounded and convex and let c be a symplectic capacity. For every $\varepsilon > 0$ there is $\delta > 0$ such that $d_H(A, B) < \delta$ with $B \subseteq \mathbb{R}^{2n}$ open and convex implies*

$$|c(A, \omega_0) - c(B, \omega_0)| < \varepsilon.$$

Proof. Let $\delta > 0$ and let $B \subseteq \mathbb{R}^{2n}$ be open and convex such that $d_H(A, B) < \delta$. This means that every point in A has Euclidean distance at most δ to B and vice versa:

$$\inf_{b\in B} ||a - b|| \leqslant \delta \ \ \forall a \in A,$$
$$\inf_{a\in A} ||a - b|| \leqslant \delta \ \ \forall b \in B.$$

Now consider the intersection of A and B. $A \cap B$ is open, convex and contained in both A and B. If we choose δ sufficiently small, we know that $A \cap B$ is nonempty because A and B are open. Furthermore, we have

$$\inf_{b \in B} ||z - b|| \leqslant \delta \ \ \forall z \in A \cap B,$$
$$\inf_{a \in A} ||z - a|| \leqslant \delta \ \ \forall z \in A \cap B.$$

In other words, both $d_H(A, A \cap B)$ and $d_H(B, A \cap B)$ are at most δ.

Next, let $u \in \mathbb{R}^{2n}$. We note that the translation $\varphi\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ given by $\varphi(x) = x + u$ is a symplectic diffeomorphism. This is easy to see since φ and φ^{-1} are both differentiable and their Jacobian is the identity matrix. Therefore, we can assume that 0 is contained in the open set $A \cap B$ without changing $c(A, \omega_0)$ or $c(B, \omega_0)$. We can now scale $A \cap B$ such that it contains A and B. More precisely, we let

$$\lambda^* = \inf\{\lambda > 0\colon A \cup B \subseteq \lambda \cdot (A \cap B)\}.$$

We observe that $\lambda^* \geqslant 1$. Moreover, for every $\varepsilon' > 0$ we can pick δ small enough such that $\lambda^* \leqslant 1 + \varepsilon'$. Let us fix some $\varepsilon > 0$. If $c(A, \omega_0) = 0$, then we choose some δ such that $A \cap B$ is nonempty. Otherwise, we choose δ such that

$$\lambda^* \leqslant \sqrt{1 + \frac{\varepsilon}{c(A, \omega_0)}}.$$

Note that $c(A, \omega_0)$ is finite since A is contained in some ball with radius R and so $c(A, \omega_0) \leqslant \pi R^2$. We have

$$c(A \cap B, \omega_0) \leqslant c(A, \omega_0) \leqslant c(\lambda^* \cdot (A \cap B), \omega_0) = (\lambda^*)^2 c(A \cap B, \omega_0),$$
$$c(A \cap B, \omega_0) \leqslant c(B, \omega_0) \leqslant c(\lambda^* \cdot (A \cap B), \omega_0) = (\lambda^*)^2 c(A \cap B, \omega_0).$$

This concludes the proof since

$$\begin{aligned} |c(A, \omega_0) - c(B, \omega_0)| &\leqslant (\lambda^*)^2 c(A \cap B, \omega_0) - c(A \cap B, \omega_0) \\ &\leqslant \left((\lambda^*)^2 - 1\right) c(A, \omega_0) \\ &\leqslant \varepsilon. \end{aligned}$$

□

In this thesis we mainly consider convex sets, most notably polytopes, instead of open sets. Therefore, it is convenient to generalize the definition of a symplectic capacity to arbitrary subsets of $\mathbb{R}^{2n}$. We do so by letting

$$c(A, \omega) := \inf\{c(B, \omega)\colon B \text{ open and } A \subseteq B\},$$

where $A \subseteq \mathbb{R}^{2n}$ is not open, c is a symplectic capacity and ω is a symplectic form. We note that all properties of symplectic capacities that we established so far also hold for non-open, bounded, convex sets, provided we consider the standard symplectic form ω_0. This is a consequence of Lemma 2.3.3 because we can approximate every bounded, convex set with open, bounded, convex sets from the outside and from the inside.

Perhaps the most important property of symplectic capacities is the fact that their existence is sufficient to prove Gromov's non-squeezing theorem [42]. In fact, this motivated symplectic capacities in the first place. The theorem states a restriction on what symplectic maps can achieve. In this regard, not much was known prior to Gromov's theorem, except for the fact that symplectic maps preserve the volume.

Theorem 2.3.4. *[42] There is an injective symplectic map $\varphi\colon B_r(0) \to Z_R(0)$ if and only if $R \geqslant r$.*

Proof. It is easy to find a suitable map φ if $R \geqslant r$, such as the identity map given by $\varphi\colon x \mapsto x$. Conversely, if we assume that there is a symplectic capacity c, we get

$$\pi r^2 = c(B_r(0), \omega_0) \leqslant c(Z_R(0), \omega_0) = \pi R^2. \qquad \square$$

2.3.1 The Ekeland-Hofer-Zehnder capacity

So far, we elaborated on symplectic capacities but we have yet to demonstrate that symplectic capacities exist at all. Many constructions have been found since Gromov stated his non-squeezing theorem. Two examples are the Ekeland-Hofer capacity [49] and the Hofer-Zehnder capacity [50]. It is known that these two symplectic capacities coincide on convex bodies, i.e. nonempty, compact, convex sets with 0 in the interior [50], [98]. Thus, the name Ekeland-Hofer-Zehnder capacity makes sense if we restrict our focus to convex bodies.

To state the definition of the Ekeland-Hofer-Zehnder capacity, we require the notion of an outer normal cone. The *outer normal cone* of a set $A \subseteq \mathbb{R}^n$ at a boundary point $y \in \partial A$ is the cone

$$N_A(y) = \{m \in \mathbb{R}^n\colon m^T(z - y) \leqslant 0 \text{ for all } z \in A\}.$$

Throughout this thesis we abbreviate the term "outer normal cone" with "normal cone". This carries some ambiguity since the term "inner normal cone" also exists. However, all normal cones appearing in this thesis are outer normal cones.

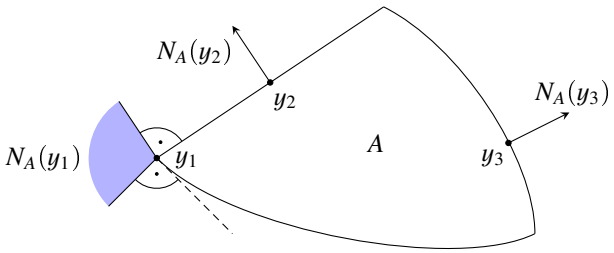


Figure 2.1: Three normal cones of a two-dimensional convex set. The normal cones at the points y_2 and y_3 are one-dimensional, indicated by arrows. The normal cone at y_1 is two-dimensional, indicated by the blue area.

Definition 2.3.5. *The Ekeland-Hofer-Zehnder capacity (short: EHZ capacity) is defined by*

$$c_{\mathrm{EHZ}}(C) = \min\{\mathbb{A}(x) \colon x \text{ is a generalized closed characteristic on } \partial C\}$$

for every convex body C. Here, a generalized closed characteristic on ∂C is a curve $x\colon [0,1] \to \partial C$ that is closed, i.e. fulfils $x(0) = x(1)$, and satisfies

$$J\dot{x}(t) \in N_C(x(t))$$

for almost every $t \in [0,1]$.

We abide by the convention that the minimum over an empty set equals infinity. The reason for the adjective "generalized" is that closed characteristics were originally defined for convex sets with smooth boundary. The usage of the normal cone to include convex sets with non-smooth boundary began later. Therefore, if we use the term "closed characteristic" we mean a generalized closed characteristic on ∂C, where C has smooth boundary.

To see that the minimum in the definition is attained we refer to [49] and [51]. In general, finding a closed characteristic that minimizes the action functional is a difficult task. An exception to this is the case where we study a convex body of the form $C = K \times T$ with convex bodies $K, T \subseteq \mathbb{R}^n$. Such a body C is sometimes called the *Lagrangian product* of K and T. We elaborate more on this case in Chapter 4. While there is not much known about the EHZ capacity in general, the situation is different for a specific class of Lagrangian products. More precisely,

$$c_{\mathrm{EHZ}}(K \times K^\circ) = 4$$

for every centrally symmetric convex body $K \subseteq \mathbb{R}^n$ [81]. Here, $K^\circ \subseteq \mathbb{R}^n$ denotes the *polar set* of K:

$$K^\circ = \{y \colon y^T z \leqslant 1 \ \ \forall\, z \in K\}.$$

Another formulation for the EHZ capacity is due to Abbondandolo and Majer [1]. In the proof one examines two functions H_C and H_{C° that are in a certain sense a dual pair. This process is reminiscent of Clarke's duality [35], a variational principle for periodic solutions of Hamiltonian systems (see Chapter 2.4.1). We state the proof for the sake of completeness.

Theorem 2.3.6. *[1] Let $C \subseteq \mathbb{R}^{2n}$ be a convex set with smooth boundary and $0 \in \operatorname{int} C$. Then:*

$$\frac{1}{4c_{\mathrm{EHZ}}(C)} = \max\left\{\mathbb{A}(\xi) \mid \xi\colon [0,1] \to \mathbb{R}^{2n} \text{ absolutely continuous,} \right.$$
$$\left. \xi(0) = \xi(1),\ \ \dot{\xi} \in C^\circ \text{ a.e.}\right\}.$$

Proof. For a set $A \subseteq \mathbb{R}^m$, $m \in \mathbb{N}$, we let μ_A be the *Minkowski functional*, i.e.:

$$\mu_A\colon \mathbb{R}^m \to \mathbb{R},\ \ \mu_A(z) = \inf\{t \in \mathbb{R}_{\geqslant 0} \colon z \in tA\} \ \ \forall\, z \in \mathbb{R}^m.$$

It is easy to see that

$$\mu_A(rz) = r\mu_A(z) \tag{2.7}$$

holds for every $r \in \mathbb{R}$ and $z \in \mathbb{R}^m$: We have $z \in tA$ if and only if $rz \in rtA$ for every $r, t \in \mathbb{R}$ with $r \neq 0$. Additionally, it is clear that (2.7) holds for $r = 0$, provided that $0 \in A$. We let

$$\begin{aligned} a_2(C) = \max \big\{ \mathbb{A}(\xi) \mid \xi \colon [0,1] \to \mathbb{R}^{2n} \text{ absolutely continuous,} \\ \xi(0) = \xi(1), \ \ ||\mu_{C^\circ}(\dot{\xi})||_2 \leqslant 1 \text{ a.e.} \big\}, \end{aligned}$$

$$\begin{aligned} a_\infty(C) = \max \big\{ \mathbb{A}(\xi) \mid \xi \colon [0,1] \to \mathbb{R}^{2n} \text{ absolutely continuous,} \\ \xi(0) = \xi(1), \ \ ||\mu_{C^\circ}(\dot{\xi})||_\infty \leqslant 1 \text{ a.e.} \big\}, \end{aligned}$$

where $||\cdot||_2$ is the L^2-norm and $||\cdot||_\infty$ is the L^∞-norm. Note that

$$\begin{aligned} ||\mu_{C^\circ}(\dot{\xi})||_\infty \leqslant 1 &\iff \mu_{C^\circ}(\dot{\xi}) \leqslant 1 \text{ almost everywhere} \\ &\iff \dot{\xi}(t) \in C^\circ \text{ almost everywhere.} \end{aligned}$$

Both $a_2(C)$ and $a_\infty(C)$ are equal to the optimal value of the maximization problem in the statement of the theorem (see [1]). We now pursue the following strategy:

(*i*) First, we prove that

$$\mathbb{A}(x) \geqslant \frac{1}{4a_\infty(C)}$$

holds for every closed characteristic x on ∂C.

(*ii*) Second, we let $\xi \colon [0,1] \to \mathbb{R}^{2n}$ with $\xi(0) = \xi(1)$ and $||\mu_{C^\circ}(\dot{\xi})||_2 \leqslant 1$ such that the maximum in the formulation of $a_2(C)$ is attained. We construct from ξ a closed characteristic x on ∂C such that

$$\mathbb{A}(x) = \frac{1}{4a_2(C)}.$$

Let us prove (*i*). We define a function

$$H_C \colon \mathbb{R}^{2n} \to \mathbb{R}, \ \ H_C(z) = \frac{\mu_C(z)^2}{2} \text{ for all } z \in \mathbb{R}^{2n}$$

and we let $x \colon [0,1] \to \partial C$ be a closed characteristic. By definition of H_C, we have

$$\nabla H_C(z) \in N_C(z)$$

for every $z \in \partial C$. Therefore, we get a closed curve $y \colon [0,b] \to \mathbb{R}^{2n}$ with period $b > 0$ by reparametrizing x such that

$$J\dot{y}(t) = \nabla H_C(y(t))$$

for almost every $t \in [0, b]$. Furthermore, we have $y(t) \in \partial C$ and therefore

$$H_C(y(t)) = \frac{\mu_C(y(t))^2}{2} = \frac{1}{2}$$

for every $t \in [0, b]$. Next, we recall Euler's theorem for homogeneous functions [94]. If a function $f\colon \mathbb{R}^m \to \mathbb{R}$ is k-homogeneous, i.e.

$$f(\alpha z) = \alpha^k f(z) \text{ for every } z \in \mathbb{R}^m \text{ and } \alpha \in \mathbb{R},$$

then we have

$$kf(z) = (\nabla f(z))^T z \text{ for every } z \in \mathbb{R}^m.$$

We can use this identity to calculate the action of y. Note that H_C is 2-homogeneous because μ_C is 1-homogeneous:

$$\begin{aligned}\mathbb{A}(y) = -\frac{1}{2}\int_0^b \dot{y}(t)^T J y(t)\,dt = \frac{1}{2}\int_0^b (J\dot{y}(t))^T y(t)\,dt = \frac{1}{2}\int_0^b \left(\nabla H_C(y(t))\right)^T y(t)\,dt \\ = \int_0^b H_C(y(t))\,dt = \int_0^b \frac{1}{2}\,dt = \frac{b}{2}.\end{aligned}$$

We know not only the action of y but also the action of x. If $g\colon [0, b] \to [0, 1]$ is the parametrization that we applied to get the curve y, then:

$$\begin{aligned}\mathbb{A}(y) = -\frac{1}{2}\int_0^b (\dot{x \circ g})(t) J (x \circ g)(t)\,dt = -\frac{1}{2}\int_0^b g'(t)\dot{x}(g(t)) J x(g(t))\,dt \\ = -\frac{1}{2}\int_0^1 \dot{x}(t) J x(t)\,dt = \mathbb{A}(x).\end{aligned}$$

Next, we define the function H_{C° analogously to H_C:

$$H_{C^\circ} : \mathbb{R}^{2n} \to \mathbb{R}, \quad H_{C^\circ}(z) = \frac{\mu_{C^\circ}(z)^2}{2} \text{ for all } z \in \mathbb{R}^{2n}.$$

We note that the following relation holds for every $v \in \mathbb{R}^{2n}$ because C is convex and contains 0:

$$\begin{aligned}\sup_{z \in \mathbb{R}^{2n}} (v^T z - H_C(z)) &= \sup_{r \geqslant 0} \sup_{\mu_C(z)=1} \left(r v^T z - \frac{\mu_C(rz)^2}{2} \right) = \sup_{r \geqslant 0} \left(r \sup_{\mu_C(z)=1} v^T z - \frac{r^2}{2} \right) \\ &= \sup_{r \geqslant 0} \left(r \mu_{C^\circ}(v) - \frac{r^2}{2} \right) = \frac{\mu_{C^\circ}(v)^2}{2} = H_{C^\circ}(v).\end{aligned} \tag{2.8}$$

Since H_{C° satisfies (2.8) we say that it is the *Fenchel conjugate* of H_C. Using the Legendre reciprocity formula (see [35]), we get

$$\nabla H_{C^\circ}\left(\nabla H_C(z)\right) = z \text{ and } \nabla H_C\left(\nabla H_{C^\circ}(z)\right) = z \tag{2.9}$$

for every $z \in \mathbb{R}^{2n}$. With Euler's theorem it follows that

$$2H_{C^\circ}(\nabla H_C(z)) = (\nabla H_C(z))^T \nabla H_{C^\circ}(\nabla H_C(z)) = (\nabla H_C(z))^T z = 2H_C(z) \tag{2.10}$$

for every $z \in \mathbb{R}^{2n}$. We now consider the closed curve

$$\gamma\colon [0,1] \to \mathbb{R}^{2n}, \;\; \gamma(t) = \frac{1}{b} J y(bt).$$

With (2.10) we see that γ fulfils the following for almost every $t \in [0,1]$:

$$H_{C^\circ}\left(\dot{\gamma}(t)\right) = H_{C^\circ}\left(J\dot{y}(bt)\right) = H_{C^\circ}\left(\nabla H_C(y(bt))\right) = H_C(y(bt)) = \frac{1}{2}.$$

In particular, it follows that $\mu_{C^\circ}\left(\dot{\gamma}(t)\right) = 1$ and hence $\dot{\gamma}(t)$ is contained in C° for almost every $t \in [0,1]$. Therefore, γ is a feasible solution in the maximization problem that defines $a_\infty(C)$. Furthermore, we can relate the action of γ to the action of x:

$$\begin{aligned}\mathbb{A}(\gamma) = -\frac{1}{2}\int_0^1 (\dot{\gamma}(t))^T J\gamma(t)\,dt &= -\frac{1}{2b}\int_0^1 (\dot{y}(bt))^T J^T J J y(bt)\,dt \\ &= -\frac{1}{2b^2}\int_0^b (\dot{y}(t))^T J y(t)\,dt = \frac{1}{b^2}\mathbb{A}(y) = \frac{1}{b^2}\mathbb{A}(x).\end{aligned}$$

Together with the fact that the action of x is $b/2$ we get:

$$\frac{1}{4a_\infty(C)} \leqslant \frac{1}{4\mathbb{A}(\gamma)} = \frac{b^2}{4\mathbb{A}(x)} = \mathbb{A}(x).$$

Now we proceed with part (ii). We let $\xi\colon [0,1] \to \mathbb{R}^{2n}$ be an absolutely continuous curve such that

$$\begin{aligned}\xi = \operatorname{argmax}\big\{\mathbb{A}(\varphi) \mid \varphi\colon [0,1] \to \mathbb{R}^{2n} \text{ absolutely continuous,} \\ \varphi(0) = \varphi(1), \;\; ||\mu_{C^\circ}(\dot{\varphi})||_2 \leqslant 1 \text{ a.e.}\big\}.\end{aligned} \tag{2.11}$$

In particular, we have $a_2(C) = \mathbb{A}(\xi)$. We observe that $||\mu_{C^\circ}(\dot{\xi})||_2 = 1$ because otherwise we can scale ξ by some factor $\theta \geqslant 1$ to get a curve $\theta\xi$ with $||\mu_{C^\circ}(\theta\dot{\xi})||_2 \leqslant 1$. Furthermore, its action is

$$\mathbb{A}(\theta\xi) = \theta^2\mathbb{A}(\xi) > \mathbb{A}(\xi),$$

which violates the definition of ξ. Note that $\mathbb{A}(\xi)$ is positive since we already found a curve γ in part (i) with

$$1 = ||\mu_{C^\circ}(\dot{\gamma})||_\infty \geqslant ||\mu_{C^\circ}(\dot{\gamma})||_2$$

and with action $\mathbb{A}(\gamma) = 1/2b > 0$. Additionally, we notice that ξ is still a maximizer of (2.11) if we add the constraint $\Phi_C(\xi) = 1/4$, where

$$\Phi_C(\xi) := \frac{1}{2}\int_0^1 H_{C^\circ}(\dot{\xi}(t))\,dt = \frac{1}{4}\int_0^1 \mu_{C^\circ}(\dot{\xi}(t))^2\,dt = \frac{1}{4}||\,\mu_{C^\circ}(\dot{\xi})||_2^2.$$

According to the Lagrange multiplier theorem [38] there is some $\lambda \in \mathbb{R}$ such that

$$\nabla\mathbb{A}(\xi) = \lambda\nabla\Phi_C(\xi). \tag{2.12}$$

We note that both $\mathbb{A}$ and Φ_C are 2-homogeneous. Therefore, we can use Euler's theorem to get

$$2a_2(C) = 2\mathbb{A}(\xi) = \nabla\mathbb{A}(\xi)[\xi] = \lambda\nabla\Phi_C(\xi)[\xi] = 2\lambda\Phi_C(\xi) = \frac{\lambda}{2}. \tag{2.13}$$

Thus, it is $\lambda = 4a_2(C)$. We note that the differentials of $\mathbb{A}$ and $\Phi_C(\xi)$ are given by

$$\begin{aligned}
\nabla\Phi_C(\xi)[\eta] := \frac{d}{dh}\Phi_C(\xi+h\eta)\Big|_{h=0} &= \frac{1}{2}\int_0^1 \dot{\eta}(t)^T\nabla H_{C^0}(\dot{\xi}(t)+h\dot{\eta}(t))\,dt\Big|_{h=0} \\
&= \frac{1}{2}\int_0^1 \dot{\eta}(t)^T\nabla H_{C^0}(\dot{\xi}(t))\,dt, \\
\nabla\mathbb{A}(\xi)[\eta] := \frac{d}{dh}\mathbb{A}(\xi+h\eta)\Big|_{h=0} &= -\frac{1}{2}\int_0^1 \dot{\xi}(t)^T J\eta(t) + \dot{\eta}(t)^T J\xi(t) + 2h\dot{\eta}(t)^T J\eta(t)\,dt\Big|_{h=0} \\
&= -\frac{1}{2}\int_0^1 \dot{\xi}(t)^T J\eta(t) + \dot{\eta}(t)^T J\xi(t)\,dt \\
&= -\int_0^1 \dot{\eta}(t)^T J\xi(t)\,dt
\end{aligned}$$

for every closed curve $\eta \in H^1([0,1],\mathbb{R}^{2n})$, where

$$H^1([0,1],\mathbb{R}^{2n}) = \left\{\eta \in L^2([0,1],\mathbb{R}^{2n})\colon \dot{\eta} \in L^2([0,1],\mathbb{R}^{2n})\right\}$$

is the *Sobolev space* of order 1 on $[0,1]$. In the last equation we use integration by parts and the fact that $\xi(0) = \xi(1)$ and $\eta(0) = \eta(1)$ holds. Combining (2.12) with (2.13) and plugging in the formulas for the derivatives, we get for every closed curve $\eta \in H^1([0,1],\mathbb{R}^{2n})$:

$$0 = \lambda\nabla\Phi_C(\xi)[\eta] - \nabla A(\xi)[\eta] = \int_0^1 \dot{\eta}(t)^T\left(2a_2(C)\nabla H_{C^0}(\dot{\xi}(t)) + J\xi(t)\right)\,dt.$$

The Du Bois-Reymond Lemma [38] now states that the function

$$2a_2(C)\nabla H_{C^\circ}(\dot{\xi}(t)) + J\xi(t)$$

is a constant almost everywhere. So, there is $u \in \mathbb{R}^{2n}$ such that for almost every $t \in [0,1]$ we have

$$-J\xi(t) + u = 2a_2(C)\nabla H_{C^\circ}(\dot{\xi}(t)).$$

We apply the 1-homogeneous function ∇H_C to both sides and use (2.9) to get

$$\nabla H_C(-J\xi(t) + u) = 2a_2(C)\nabla H_C(\nabla H_{C^\circ}(\dot{\xi}(t))) = 2a_2(C)\dot{\xi}(t).$$

In particular, if we let $y\colon [0,1] \to \mathbb{R}^{2n}$ be the closed curve defined by $y(t) = -J\xi(t) + u$, then:

$$J\dot{y}(t) = \frac{1}{2a_2(C)}\nabla H_C(y(t)) \tag{2.14}$$

for almost every $t \in [0,1]$. Thus, the curve y lives on a level set of $H_C(y(t))$ (see Figure 2.2). More precisely, there is $E > 0$ such that $H_C(y(t)) = E$ for all $t \in [0,1]$. Now the closed curve $x\colon [0,1] \to \mathbb{R}^{2n}$ that we define by $x(t) = y(t)/\sqrt{2E}$ fulfils:

$$H_C(x(t)) = \frac{1}{2E}H_C(y(t)) = \frac{1}{2}$$

for every $t \in [0,1]$. By definition of H_C, it follows that $x(t) \in \partial C$ and together with (2.14) we get that x is a closed characteristic on ∂C. To complete part (ii) and hence the proof of Theorem 2.3.6, we calculate the action of x. To this end, we combine our previous considerations about the curves x and y and we use Euler's theorem:

$$\begin{aligned}\mathbb{A}(x) = \frac{1}{2E}\mathbb{A}(y) &= \frac{1}{4E}\int_0^1 (J\dot{y}(t))^T y(t)\,dt = \frac{1}{8Ea_2(C)}\int_0^1 \nabla H_C(y(t))^T y(t)\,dt \\ &= \frac{1}{4Ea_2(C)}\int_0^1 H_C(y(t))\,dt = \frac{E}{4Ea_2(C)} = \frac{1}{4a_2(C)}.\end{aligned}$$

□

We note that the requirement $0 \in \operatorname{int} C$ has a rather technical reason. It ensures that the polar set fulfils some necessary properties. However, we can also apply Theorem 2.3.6 to every nonempty convex set that does not contain 0 in its interior by shifting it. The corresponding map that performs the shift is a symplectic diffeomorphism and thus preserves the EHZ capacity.

Due to a result from Haim-Kislev, a similar version of Theorem 2.3.6 also holds for polytopes [47]. We pursue this approach in greater detail in Chapter 5.

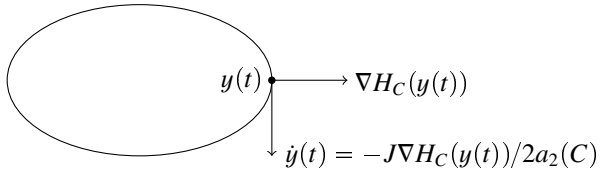


Figure 2.2: A level curve of H_C that contains $y(t)$, which is a solution of (2.14). We observe that y moves along the level curve.

2.4 Applications and related problems

2.4.1 Hamiltonian mechanics

Symplectic geometry has a tight connection to Hamiltonian mechanics, a specific formulation of classical mechanics. It is one of the first applications that comes to mind because symplectic geometry was developed as a mathematical formalization of Hamiltonian mechanics. To make this connection clear, we consider a physical system in which we investigate the time evolution of some particle. As an example, we want to know how a celestial object like a planet moves. To determine this, we need to know the current position $p(0)$ of the planet and its momentum $q(0)$. Thus, we work in the space $\mathbb{R}^6 = \mathbb{R}^3 \times \mathbb{R}^3$ or more generally in the space $\mathbb{R}^{2n}$. Note that we write variables in the form (p, q) for an easier distinction between the position p and the momentum q. According to Hamiltonian mechanics, we can determine the motion of the planet by solving the following differential equation:

$$\frac{dp}{dt} = \frac{\partial H}{\partial q}, \quad \frac{dq}{dt} = -\frac{\partial H}{\partial p}, \tag{2.15}$$

where $H = H(p, q)$ is a differentiable, real-valued function that describes the physical system. H is called the *Hamiltonian*. A more compact way to write (2.15) is

$$\dot{x}(t) = -J\nabla H(x(t)), \tag{2.16}$$

where we let $x(t) = (p(t), q(t))$ for every $t \geqslant 0$. We define the *Hamiltonian vector field* $X_H \colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, that is associated to (2.16), by

$$X_H(x) = -J\nabla H(x).$$

Given an initial value x_0, the differential equation (2.16) becomes

$$\begin{aligned} \frac{d}{dt}\varphi^t(x_0) &= X_H(\varphi^t(x_0)) \ \ \forall\, t > 0, \\ \varphi^0(x_0) &= x_0, \end{aligned} \tag{2.17}$$

if we let $x(t) = \varphi^t(x_0)$. The function φ^t is called a flow of the vector field X_H. It maps the initial position and the initial momentum of the planet to its position and momentum at time t. It is well known that every flow of X_H is symplectic. Hence, investigating the

motion of some object amounts to the study of symplectic maps. There are several ways to prove this. We present the strategy used in [74] since it requires less specific background knowledge.

Theorem 2.4.1. *[74] Let $H\colon \mathbb{R}^{2n} \to \mathbb{R}$ be twice differentiable and let φ^t be a flow of the vector field X_H. Then φ^t is symplectic.*

Proof. By definition of a flow, we have

$$\frac{d}{dt}\varphi^t(x_0) = X_H(\varphi^t(x_0)) = -J\nabla H(\varphi^t(x_0))$$

for every initial value $x_0 \in \mathbb{R}^{2n}$. We calculate the Jacobian at x_0 for each side of this equation, using the chain rule, to find

$$D_{\frac{d}{dt}\varphi^t}(x_0) = \frac{d}{dt}D_{\varphi^t}(x_0) = -J\nabla^2 H(\varphi^t(x_0))D_{\varphi^t}(x_0).$$

Here, the differential $d/dt\, D_{\varphi^t}(x_0)$ is to be understood componentwise. To make this equation clearer, we let $\Phi(t)$ be the Jacobian of φ^t at x_0 and $S(t)$ be the Hessian of H at $\varphi^t(x_0)$. Note that $S(t)$ is a symmetric matrix. Now the equation takes the form

$$\frac{d}{dt}J\Phi(t) = S(t)\Phi(t),$$

where we use the fact that $J^2 = -I$. This implies:

$$\begin{aligned}
\frac{d}{dt}\Phi(t)^T J\Phi(t) &= \Phi(t)^T J\left(\frac{d}{dt}\Phi(t)\right) + \left(\frac{d}{dt}\Phi(t)^T\right) J\Phi(t)\\
&= \Phi(t)^T S\Phi(t) + \left(\Phi(t)^T J^T\left(\frac{d}{dt}\Phi(t)\right)\right)^T\\
&= \Phi(t)^T S\Phi(t) - \left(\Phi(t)^T S\Phi(t)\right)^T\\
&= \Phi(t)^T(S - S^T)\Phi(t) = 0.
\end{aligned}$$

Furthermore, we have that φ^0 is the identity map according to (2.17). Thus, it is $\Phi(0) = I$ and

$$\Phi(0)^T J\Phi(0) = J.$$

Together, this implies that $\Phi(t)^T J\Phi(t) = J$ holds for every $t \geqslant 0$. In particular, we have

$$\omega_0(u, v) = u^T Jv = u^T\Phi(t)^T J\Phi(t)v = \omega_0(D_{\varphi^t}(x_0)u, D_{\varphi^t}(x_0)v)$$

for every $u, v, x_0 \in \mathbb{R}^{2n}$. □

We can deduce an interpretation of Gromov's non-squeezing theorem from Theorem 2.4.1. Let us assume the exact position and momentum of the object that we are studying is not known. Instead, we have an estimate of position and momentum in the form of a region of possible positions and momenta. This region may be a Euclidean ball of radius r in $\mathbb{R}^{2n}$. We can apply Gromov's non-squeezing theorem to find that the region of possible positions and momenta of the object fits at no time into an infinitely long cylinder $Z_R(0)$ wit radius $R < r$.

2.4.2 The symmetric Mahler conjecture and Viterbo's conjecture

Symplectic geometry and in particular the EHZ capacity have a striking connection to the following old problem in convex geometry. For a convex body $K \subseteq \mathbb{R}^n$ and $z \in \mathbb{R}^n$ we let

$$K^z = \left\{ y \in \mathbb{R}^n \colon (y-z)^T (x-z) \leqslant 1 \ \ \forall\, x \in K \right\}.$$

The task is to determine the *volume product*, which is defined by

$$\mathcal{P}(K) := \min_{z \in \operatorname{int} K} \operatorname{vol} K \cdot \operatorname{vol} K^z.$$

In [88] it is shown that the minimum is attained at a unique point, which is called the Santaló point of K. Furthermore, if K is centrally symmetric, i.e. $K = -K$, then the Santaló point is the origin. Note that K^z is the polar set K° of K for $z = 0$. Mahler studied the volume product $\mathcal{P}(K)$ for centrally symmetric, convex sets K in [72] and formulated the following conjecture, which is known as the symmetric Mahler conjecture.

Conjecture 2.4.2. *[72] Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric, convex set. Then:*

$$\mathcal{P}(K) = \operatorname{vol} K \cdot \operatorname{vol} K^\circ \geqslant \frac{4^n}{n!}.$$

It is easy to see that Mahler's conjecture holds for $n = 1$. In this case K is an interval $[-s, s]$ for some $s > 0$. Its one-dimensional volume is $2s$. The polar set of K is $[-1/s, 1/s]$ and has volume $2/s$. Thus, we have $\mathcal{P}(K) = 4$. Moreover, it is known that the symmetric Mahler conjecture is true for $n = 2$ [71] and $n = 3$ [55]. The problem is open for every $n \geqslant 4$.

The EHZ capacity relates the symmetric Mahler conjecture to a conjecture that was posed more recently by Viterbo [99]. Unlike Mahler's conjecture, Viterbo's conjecture is of symplectic nature.

Conjecture 2.4.3. *[99] Let c be a symplectic capacity and let $C \subseteq \mathbb{R}^{2n}$ be a convex set. Then:*

$$\frac{c(C, \omega_0)}{c(B_1(0), \omega_0)} \leqslant \left(\frac{\operatorname{vol} C}{\operatorname{vol} B_1(0)} \right)^{1/n}.$$

Let us consider a special case of Viterbo's conjecture. First, we assume that C has the form $C = K \times K^\circ$ for some centrally symmetric, convex body $K \subseteq \mathbb{R}^n$. Second, we consider the EHZ capacity. Third, we note that we know the EHZ capacity of C due to the following result.

Theorem 2.4.4. *[10],[81] Let $K \subseteq \mathbb{R}^n$ be a centrally symmetric, convex set. Then:*

$$c_{\mathrm{EHZ}}(K \times K^\circ, \omega_0) = 4.$$

The proof of this theorem uses a relation between the EHZ capacity and the notion of closed Minkowski billiards. We study this relation thouroughly in Chapter 4 but we omit the proof of Theorem 2.4.4.

If we assume that Viterbo's conjecture holds, then we get:

$$\frac{4^n}{\pi^n} = \left(\frac{c_{\mathrm{EHZ}}(K \times K^\circ, \omega_0)}{c_{\mathrm{EHZ}}(B_1(0), \omega_0)}\right)^n \leqslant \frac{\operatorname{vol} K \times K^\circ}{\operatorname{vol} B_1(0)} = \frac{\operatorname{vol} K \cdot \operatorname{vol} K^\circ}{\pi^n/n!}.$$

Thus, Viterbo's conjecture implies the symmetric Mahler conjecture. Conversely, the symmetric Mahler conjecture implies Viterbo's conjecture for the special case $c = c_{\mathrm{EHZ}}$ and $C = K \times K^\circ$ with centrally symmetric, convex K:

$$\left(\frac{c_{\mathrm{EHZ}}(K \times K^\circ, \omega_0)}{c_{\mathrm{EHZ}}(B_1(0), \omega_0)}\right)^n = \frac{4^n}{\pi^n} \leqslant \frac{\operatorname{vol} K \cdot \operatorname{vol} K^\circ}{\pi^n/n!} = \frac{\operatorname{vol} K \times K^\circ}{\operatorname{vol} B_1(0)}.$$

For a more extensive literature about the volume product and beyond its connection to symplectic geometry, we refer to [21], [22], [62] and [67].

Chapter Three

Preliminaries and optimization techniques

The formulations for the EHZ capacity in Definition 2.3.5 and Theorem 2.3.6 are minimization and maximization problems. Thus, being familiar with optimization techniques is beneficial for the computation of the EHZ capacity. This chapter introduces some important optimization concepts in two steps. First, we take a look at a specific, well-established optimization problem, the quadratic assignment problem. As we discuss later in Chapter 5, the calculation of the EHZ capacity has a relation to this combinatorial optimization problem. Afterwards, we introduce conic optimization, which is a meaningful and wide class of optimization problems. Furthermore, we examine some important instances in greater detail.

3.1 The quadratic assignment problem

For an extensive exposition on the quadratic assignment problem see [26].

Imagine our task is to build n facilities denoted with the numbers $1, \dots, n$. For this we are given n locations which we also denote with the numbers $1, \dots, n$. To each of these locations we need to assign exactly one facility. For each pair of indices $1 \leqslant i, j \leqslant n$ we know the distance between the locations i and j and denote it with $d_{i,j}$. Moreover, the n facilities ship goods among each other. We denote with $f_{i,j}$ the amount of goods that are shiped from facility i to facility j. We can view the quantity $f_{i,j}d_{k,l}$ as the cost of exchanging goods between the facilities i and j if we build facility i at location k and facility j at location l. Additionally, building a facility i at a location k causes building costs $b_{i,k}$. This raises the question: What assignment has minimal cost? In other words we want to solve the following minimization problem:

$$\min_{\sigma \in \mathrm{Sym}_n} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{i,j} d_{\sigma(i),\sigma(j)} + \sum_{i=1}^{n} b_{i,\sigma(i)},$$

where Sym_n denotes the symmetric group of degree n. This problem is called the *quadratic assignment problem* (QAP) and was first formulated in 1957 [63]. We can write the QAP in

a neater way using matrix notation. The input is given by the three matrices $F, D, B \in \mathbb{R}^{n\times n}$ with

$$F_{i,j} = f_{i,j},\ D_{i,j} = d_{i,j},\ B_{i,j} = b_{i,j}$$

for all $i, j \in \{1, \dots, n\}$. Furthermore, we define the *permutation matrix* $X^\sigma \in \mathbb{R}^{n\times n}$ for a permutation $\sigma \in \mathrm{Sym}_n$ by

$$(X^\sigma)_{i,j} = \begin{cases} 1, & \text{if } i = \sigma(j), \\ 0, & \text{otherwise.} \end{cases}$$

Let $\Pi_n = \{X^\sigma : \sigma \in \mathrm{Sym}_n\}$ be the set of $(n \times n)$-permutation matrices. We can now express the QAP as

$$\min_{X\in\Pi_n} \left(\left\langle F, X^T D X\right\rangle + \langle B, X\rangle\right). \tag{3.1}$$

Throughout this thesis $\langle\cdot,\cdot\rangle$ denotes the *trace inner product*, i.e. for two matrices $X, Y \in \mathbb{R}^{n\times n}$ we have

$$\langle X, Y\rangle = \mathrm{tr}(X^T Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} X_{i,j} Y_{i,j}.$$

It is easy to see that the traveling salesman problem, which we can write in the form

$$\min_{\sigma\in\mathrm{Sym}_n} \sum_{i=1}^{n} d_{\sigma(i),\sigma(i+1)},$$

is a special case of QAP. One just sets $b_{i,j} = 0$ and $f_{i,j} = \delta_{j,i+1}$. Therefore, QAP is at least as difficult as the traveling salesman problem, which is NP-complete [82]. Furthermore, the traveling salesman problem cannot be approximated within some constant factor unless P = NP (see [64] or [87]). Thus, the same is true for QAP.

Even though QAP was formally stated more than 50 years ago, it is still studied actively. Thus, there is a variety of techniques to approach this problem. We can subdivide these techniques into two classes: Those which derive an exact solution and those which find an approximation. In general, solving QAP exactly for $n > 20$ is very difficult [14].

Here, we present an approximative approach by Finke, Burkard and Rendl [37] that relies on the formulation (3.1) for the case $B = 0$. It utilizes the fact that there is a convenient representation of the quadratic term $\left\langle F, X^T D X\right\rangle$ via the eigenvalues of F and D if both F and D are symmetric.

Theorem 3.1.1. *[37] Let $Y_1, Y_2 \in \mathbb{R}^{n\times n}$ be symmetric matrices. Let $\lambda \in \mathbb{R}^n$ be the vector of eigenvalues of Y_1 and let $x_1, \dots, x_n$ be the corresponding orthonormal eigenvectors. Similarly, let $\mu \in \mathbb{R}^n$ be the vector of eigenvalues of Y_2 and let $y_1, \dots, y_n$ be the corresponding orthonormal eigenvectors. Then:*

(i) $\quad \langle Y_1, Y_2\rangle = \lambda^T S \mu,$
where $S = \left((x_i^T y_j)^2\right)_{i,j\in\{1,\dots,n\}}$ is a doubly stochastic matrix.

(ii) $\quad \min_{X\in\Pi_n} \lambda^T X \mu \leqslant \langle Y_1, Y_2\rangle \leqslant \max_{X\in\Pi_n} \lambda^T X \mu.$

Here, a *doubly stochastic matrix* is a matrix with nonnegative entries whose row and column sums are all equal to 1. If we assume that F and D are symmetric, we can plug in F for Y_1 and X^TDX for Y_2 to receive an upper and a lower bound on $\langle F, X^TDX \rangle$ that only depends on the eigenvalues of F and X^TDX. Since all matrices of the form X^TDX with $X \in \Pi_n$ have the same eigenvalues as D, it is sufficient to compute the eigenvalues $\lambda_1 \leqslant \ldots \leqslant \lambda_n$ of F and $\mu_1 \leqslant \ldots \leqslant \mu_n$ of D and to calculate

$$\max_{X \in \Pi_n} \lambda^T X \mu = \max_{\sigma \in \mathrm{Sym}_n} \sum_{i=1}^{n} \lambda_i \mu_{\sigma(i)} = \sum_{i=1}^{n} \lambda_i \mu_i \quad \text{and}$$

$$\min_{X \in \Pi_n} \lambda^T X \mu = \min_{\sigma \in \mathrm{Sym}_n} \sum_{i=1}^{n} \lambda_i \mu_{\sigma(i)} = \sum_{i=1}^{n} \lambda_i \mu_{n-i+1}.$$

The situation that we are facing in Chapter 5 is slightly different in the sense that the input matrices F and D are both nonsymmetric. However, we can make them skew-symmetric while maintaining the form of a QAP. For this reason, we omit the proof of Theorem 3.1.1 here. Instead, we carry the proof of Finke, Burkard and Rendl over to the skew-symmetric case in Chapter 5.

3.2 Conic optimization

We now turn our attention to a wide class of optimization problems, namely conic optimization. A more prominent term for the same class is convex optimization, where one seeks to optimize a convex function over a convex set. The defining idea of conic optimization is to formulate convex optimization problems using convex cones to gain convenient expressions [80]. We introduce the concept of conic optimization by following [16]. Let us start with some basic notions.

Definition 3.2.1. *Let V be a vector space over $\mathbb{R}$.*

(i) A (convex) cone $K \subseteq V$ is a set that satisfies $\alpha x + \beta y \in K$ for every $x, y \in K$ and every $\alpha, \beta \geqslant 0$.

(ii) A cone K is called pointed if it contains no straight line that passes through the origin. This means that $x, -x \in K$ implies $x = 0$.

(iii) We say that a cone is proper if it is closed, pointed and has nonempty interior.

With every cone $K \subseteq V$ we can associate a preorder $\leqslant_K$ on V by letting $x \leqslant_K y$ if and only if $y - x \in K$. We observe that this preorder is a partial order if and only if K is pointed.

A particularly powerful tool of conic optimization is its duality theory. Loosely speaking, this means that conic optimization problems come in closely related pairs that provide a systematic way to prove optimality or to give bounds on their optima. To define this concept formally, we call a nondegenerate bilinear form $(\cdot, \cdot) \colon V \times W \to \mathbb{R}$ with vector spaces V and W over $\mathbb{R}$ a *duality*. Dualities allow us to formulate corresponding pairs of cones and linear maps.

Definition 3.2.2.

(i) Let V_1, V_2, W_1, W_2 be vector spaces over $\mathbb{R}$ and let $(\cdot,\cdot)_i \colon V_i \times W_i \to \mathbb{R}$ be a duality for $i \in \{1,2\}$. Let $A \colon V_1 \to V_2$ and $A^* \colon W_2 \to W_1$ be linear maps. We say that A^* is the adjoint of A if

$$(A(e), f)_2 = (e, A^*(f))_1 \tag{3.2}$$

for every $e \in V_1,\ f \in W_2$.

(ii) Let V, W be vector spaces over $\mathbb{R}$ and let $(\cdot,\cdot) \colon V \times W \to \mathbb{R}$ be a duality. Let $K \subseteq V$ be a cone. The cone

$$K^* = \{w \in W \colon\ (v,w) \geqslant 0 \text{ for all } v \in K\}$$

is called the dual cone of K. Moreover, if $K^* = K$ we say that K is self-dual.

It is straightforward to show that A^* is unique if it exists, due to nondegeneracy of the duality $(\cdot,\cdot)_1$ in (3.2). We are now ready to state the aforementioned pair of conic programs. Like in Definition 3.2.2, we let V_1, V_2, W_1, W_2 be vector spaces over $\mathbb{R}$ and $(\cdot,\cdot)_1, (\cdot,\cdot)_2$ be corresponding dualities. Furthermore, let $K_1 \subseteq W_1$ and $K_2 \subseteq W_2$ be cones. Additionally, we let $A \colon W_2 \to W_1$ be a linear map, $b \in W_1$ and $c \in V_2$. The following optimization problem is called a *primal conic program*:

$$\begin{aligned} p^* = \sup\ & (c,x)_2 \\ & A(x) \leqslant_{K_1} b \\ & x \geqslant_{K_2} 0. \end{aligned}$$

The corresponding *dual conic program* is given by:

$$\begin{aligned} d^* = \inf\ & (y,b)_1 \\ & A^*(y) \geqslant_{K_2^*} c \\ & y \geqslant_{K_1^*} 0. \end{aligned}$$

In this thesis we only consider conic programs where $K_1 = \{0\}$ and K_2 is a proper cone. This is the most common type of conic programs in the literature. The resulting primal and dual programs are said to be in standard form and since $K_1^* = V_1$ the problems read:

$$\begin{aligned} p^* = \sup\ & (c,x)_2 \\ & A(x) = b \\ & x \geqslant_{K_2} 0 \end{aligned} \qquad \text{and} \qquad \begin{aligned} d^* = \inf\ & (y,b)_1 \\ & A^*(y) \geqslant_{K_2^*} c \\ & y \in V_1. \end{aligned} \tag{3.3}$$

We say that $x \in W_2$ is *feasible* for the primal problem if it satisfies the constraints above, i.e. if $x \in K_2$ and $A(x) = b$. If additionally $x \in \operatorname{int} K_2$ holds, we say that x is *strictly feasible*. A feasible vector x with $p^* = (c,x)_2$ is called *optimal*. Similarly, $y \in V_1$ is feasible for the dual problem if $A^*(y) - c \in K_2^*$, strictly feasible if $A^*(y) - c \in \operatorname{int} K_2^*$ and optimal if it is feasible and $d^* = (y,b)_1$.

The reason why duality is such a powerful tool is the fact that feasible and optimal solutions of the primal program provide us with information about the dual program and vice versa.

Theorem 3.2.3. *[16] Consider the pair of conic programs* (3.3)*.*

(i) Weak duality: If x is a feasible solution for the primal program and y is a feasible solution for the dual program, then:

$$(c, x)_2 \leqslant (y, b)_1 .$$

In particular, we have $p^* \leqslant d^*$.

(ii) Optimality criterion: If x is a feasible solution for the primal program and y is a feasible solution for the dual program such that

$$(A^*(y) - c, x)_2 = 0 \text{ and } (y, A(x) - b)_1 = 0,$$

then both x and y are optimal and we have $p^* = d^*$.

(iii) Complementary slackness: If x is an optimal solution for the primal program, y an optimal solution for the dual program and if $p^* = d^*$, then:

$$(A^*(y) - c, x)_2 = 0 \text{ and } (y, A(x) - b)_1 = 0.$$

(iv) Strong duality: If $p^* < \infty$ and if there is a strictly feasible solution of the primal program, then the infimum in (3.3) *is attained and $p^* = d^*$.*

Conversely, if $d^* > -\infty$ and if there is a strictly feasible solution of the dual program, then the supremum in (3.3) *is attained and $p^* = d^*$.*

The condition in Theorem 3.2.3 (iv), i.e. the existence of a strictly feasible solution for the primal or dual program, is called *Slater's condition*. For a more detailed introduction to conic programming and its duality theory see [17] and [68].

We proceed to discuss four important instances of conic optimization that appear later in this thesis.

3.2.1 Linear programming

One of the most obvious and in terms of notation simplest proper cones is the nonnegative orthant $\mathbb{R}^n_{\geqslant 0}$. Thus, it is not surprising that this cone leads to a classical type of conic programs that is well understood by now, namely linear programs [89]. First, let us specify the vector spaces, the cone and the dualities in (3.3):

$$\begin{aligned} V_1 = W_1 &= \mathbb{R}^m, \\ V_2 = W_2 &= \mathbb{R}^n, \\ K_2 &= \mathbb{R}^n_{\geqslant 0}, \\ (u, v)_1 &= u^T v \text{ for all } u, v \in \mathbb{R}^m, \\ (u, v)_2 &= u^T v \text{ for all } u, v \in \mathbb{R}^n. \end{aligned}$$

Furthermore, we can write the linear map $A\colon \mathbb{R}^n \to \mathbb{R}^m$ as an $(m \times n)$-matrix. Its adjoint is $A^* = A^T$. It is clear from the definition that the cone $\mathbb{R}^n_{\geqslant 0}$ is self-dual. Plugging this into the standard primal and dual conic programs (3.3) we get

$$\begin{aligned} p^* = \sup\, & c^T x \\ & Ax = b \\ & x \in \mathbb{R}^n_{\geqslant 0} \end{aligned} \qquad \text{and} \qquad \begin{aligned} d^* = \inf\, & y^T b \\ & A^T y \geqslant c \\ & y \in \mathbb{R}^m. \end{aligned} \tag{3.4}$$

This special case of conic programming is called *linear programming*. We use the abbreviation LP for both linear programming and linear program depending on the context. The same principle applies for abbreviations for the other upcoming special cases of conic programming.

LP stands out for two reasons. First, strong duality holds for LP regardless of whether there are strictly feasible solutions.

Theorem 3.2.4. *[16] Consider a pair of LPs as* (3.4)*. If there is a feasible solution for the primal LP, then $p^* = d^*$ holds. Additionally, if $p^* < \infty$, there is an optimal solution for the primal and an optimal solution for the dual LP.*

In particular, we can exchange supremum and infimum in (3.4) by maximum and minimum if we add the convention that minima and maxima can be $\pm\infty$.

Second, it is well known that solving LPs up to arbitrary precision is possible in polynomial time. As a matter of fact, there are several established algorithms that compute solutions of LPs. Schrijver gives a broad overview of such algorithms [89].

3.2.2 Second order cone programming

Another established choice for the cone K_2 is the Lorentz or second-order cone. It is exactly what people have in mind when they use the term cone in the common use of language, for example if they buy ice cream. Thus, the Lorentz cone is sometimes called ice cream cone. We define it by

$$\mathcal{L}^{n+1} = \left\{(x,t) \in \mathbb{R}^{n+1} \colon ||x|| \leqslant t\right\},$$

where $||\cdot||$ is the Euclidean norm. As before, we specify the setting:

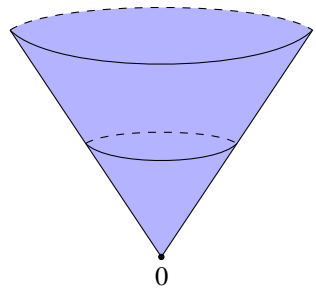


Figure 3.1: The Lorentz cone $\mathcal{L}^3$.

$$\begin{aligned} V_1 = W_1 &= \mathbb{R}^m, \\ V_2 = W_2 &= \mathbb{R}^{n+1}, \\ K_2 &= \mathcal{L}^{n+1}, \\ (u,v)_1 &= u^T v \ \text{ for all } u, v \in \mathbb{R}^m, \\ ((u,s),(v,t))_2 &= u^T v + st \ \text{ for all } (u,s),(v,t) \in \mathcal{L}^{n+1}. \end{aligned}$$

As in the previous case, we can write the linear map $A\colon \mathbb{R}^{n+1} \to \mathbb{R}^m$ as an m by $(n+1)$ matrix. We denote the ith row of A by (a_i^T, r_i) and its adjoint by A^T. Using the Cauchy-Schwarz inequality it is easy to prove that the Lorentz cone is self-dual. Thus, (3.3) becomes:

$$\begin{aligned} p^* = \sup\, & c^T x + st & d^* = \inf\, & y^T b \\ & a_i^T x + r_i t = b_i \ \ \forall\ 1 \leqslant i \leqslant m & & \left(\left[\sum_{i=1}^m y_i a_i \right] - c, \left[\sum_{i=1}^m y_i r_i \right] - s \right) \in \mathcal{L}^{n+1} \\ & (x,t) \in \mathcal{L}^{n+1}, & & y \in \mathbb{R}^m. \end{aligned}$$

This type of conic programming is called *second-order cone programming* (SOCP). We observe that LP is a special case of SOCP by picking $r_1 = \ldots = r_m = s = 0$. In fact, this caused SOCP to gain attention in 1984 when Karmakar gave a polynomial-time interior-point algorithm for LP [60]. This algorithm carries over to the case where the objective function and/or the constraints are quadratic and convex instead of linear. In the following years this was studied intensively (see for example [57],[75],[76],[79]), leading to efficient algorithms to solve SOCPs.

We note that in the primal SOCP we only consider one variable $(x,t) \in \mathcal{L}^{n+1}$. In practice however, it is often neccessary to employ multiple such variables. For this reason, the term SOCP can also refer to a conic program where we optimize over a direct product of Lorentz cones:

$$K_2 = \mathcal{L}^{n_1+1} \times \ldots \times \mathcal{L}^{n_k+1},$$

where $n_1, \ldots, n_k \in \mathbb{N}$. We remark that K_2 is again a cone. It is easy to generalize the above primal-dual pair of SOCP in standard form from the case $k = 1$ to arbitrary $k \in \mathbb{N}$:

$$\begin{aligned} p^* = \sup\, & \sum_{j=1}^k c_j^T x_j + s_j t_j \\ & \sum_{j=1}^k (a_j)_i^T x_j + (r_j)_i t_j = b_i \ \ \forall\ 1 \leqslant i \leqslant m \\ & (x_j, t_j) \in \mathcal{L}^{n_j+1} \ \ \forall\ 1 \leqslant j \leqslant k, \end{aligned}$$

$$\begin{aligned} d^* = \inf\, & y^T b \\ & \left(\left[\sum_{i=1}^m y_i (a_j)_i \right] - c_j, \left[\sum_{i=1}^m y_i (r_j)_i \right] - s_j \right) \in \mathcal{L}^{n_j+1} \ \ \forall\ 1 \leqslant j \leqslant k \\ & y \in \mathbb{R}^m. \end{aligned}$$

3.2.3 Semidefinite programming

The next cone that we consider is the cone of positive semidefinite matrices. For this, we denote the space of symmetric $n \times n$-matrices by $\mathcal{S}^n$. We say that a matrix $X \in \mathcal{S}^n$ is *positive*

semidefinite, if we have

$$y^T X y \geqslant 0$$

for every $y \in \mathbb{R}^n$. If this inequality is strict for every $y \neq 0$, then we say that X is *positive definite*. We use the notation $X \succeq 0$ (respectively $X \succ 0$) to express that a matrix X is positive semidefinite (respectively positive definite). We observe that if $X, Y \in \mathcal{S}^n$ are positive semidefinite, then so is $\alpha X + \beta Y$ for all real numbers $\alpha, \beta \geqslant 0$. Thus, the set

$$\mathcal{S}^n_{\succeq 0} := \{X \in \mathcal{S}^n \colon X \succeq 0\}$$

is a cone. Before we state the corresponding version of (3.3), we recall some properties of $\mathcal{S}^n_{\succeq 0}$.

Theorem 3.2.5. *[54] Let $X \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The following properties are equivalent:*

(i) X is positive semidefinite.

(ii) There are $u_1, \ldots, u_n \in \mathbb{R}^n$ and $\lambda_1, \ldots, \lambda_n \geqslant 0$ such that $X = \sum_{i=1}^{n} \lambda_i u_i u_i^T$.

(iii) X has a Cholesky decomposition, i.e. $X = LL^T$ for some $L \in \mathbb{R}^{n\times n}$.

(iv) The eigenvalues of X are nonnegative.

(v) X is a Gram matrix, i.e. there are $x_1, \ldots, x_n \in \mathbb{R}^n$ such that $X_{i,j} = x_i^T x_j$ for all $i, j \in \{1, \ldots, n\}$.

(vi) All principal minors of X are nonnegative. This means that for every index set $I \subseteq \{1, \ldots, n\}$:

$$\det (X_{i,j})_{i,j\in I} \geqslant 0.$$

Aside from positive semidefinite matrices we are also interested in positive definite matrices here. The reason for this is that the set of positive definite matrices is the interior of the cone of semidefinite matrices. In other words, the boundary of $\mathcal{S}^n_{\succeq 0}$ consists of positive semidefinite matrices with rank smaller than n. Moreover, every extreme ray of $\mathcal{S}^n_{\succeq 0}$ is spanned by a positive semidefinite matrix with rank 1. Conversely, every positive semidefinite matrix with rank 1 spans an extreme ray of $\mathcal{S}^n_{\succeq 0}$. We state a counterpart of Theorem 3.2.5 for positive definite matrices.

Theorem 3.2.6. *[54] Let $X \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The following properties are equivalent:*

(i) X is positive definite.

(ii) There are $u_1, \ldots, u_n \in \mathbb{R}^n$ and $\lambda_1, \ldots, \lambda_n > 0$ such that $X = \sum_{i=1}^{n} \lambda_i u_i u_i^T$.

(iii) $X = LL^T$ for some regular matrix $L \in \mathbb{R}^{n\times n}$.

(iv) The eigenvalues of X are positive.

(v) There are linearly independent vectors $x_1, \ldots, x_n \in \mathbb{R}^n$ such that $X_{i,j} = x_i^T x_j$ for all $i, j \in \{1, \ldots, n\}$.

(vi) All leading principal minors of X are positive. This means that for every index set $I = \{1, \ldots, k\}$ with $1 \leqslant k \leqslant n$:

$$\det (X_{i,j})_{i,j \in I} > 0.$$

For more properties and details of positive definite and positive semidefinite matrices, see [54]. Let us specify the corresponding setting to formulate a primal-dual pair of conic programs:

$$\begin{aligned} V_1 = W_1 &= \mathbb{R}^m, \\ V_2 = W_2 &= \mathcal{S}^n \\ K_2 &= \mathcal{S}^n_{\geq 0}, \\ (u, v)_1 &= u^T v \ \text{ for all } u, v \in \mathbb{R}^m, \\ (X, Y)_2 &= \langle X, Y \rangle := \operatorname{tr}(X^T Y) \ \text{ for all } X, Y \in \mathcal{S}^n. \end{aligned}$$

Unlike in the two previous cases, it is rather confusing to write a linear map $A \colon \mathcal{S}^n \to \mathbb{R}^m$ as a single matrix. Instead we use the following notation:

$$A(X) = \left(\langle A_i, X \rangle \right)_{i \in \{1, \ldots, m\}},$$

where $A_1, \ldots, A_m$ are $(n \times n)$-matrices. Its adjoint is given by

$$A^*(y) = \sum_{i=1}^m y_i A_i.$$

Similar to the previous cases, $\mathcal{S}^n_{\geq 0}$ is self-dual (see for instance [23]). So, (3.3) becomes

$$\begin{array}{lll} p^* = \sup \langle C, X \rangle & & d^* = \inf y^T b \\ \langle A_i, X \rangle = b_i \ \ \forall\ 1 \leqslant i \leqslant m & \text{and} & \displaystyle\sum_{i=1}^m y_i A_i - C \succeq 0 \\ X \succeq 0 & & y \in \mathbb{R}^m. \end{array} \tag{3.5}$$

Conic programs of this form are called *semidefinite programs* (SDP). By letting the input matrices $C, A_1, \ldots, A_m$ be diagonal matrices, we observe that we can write every LP as an SDP. Moreover, one can show that SOCP is a special case of SDP as well. To this end we consider a block matrix of the form

$$X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix},$$

where the matrix A is regular. The matrix $C - B^T A^{-1} B$ is called the *Schur complement* of X.

Theorem 3.2.7. *[28] Let X be a symmetric matrix having the block form*

$$X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix},$$

where A is a regular matrix. Then X is positive semidefinite if and only if A is positive definite and the Schur complement $C - B^T A^{-1} B$ is positive semidefinite.

Thus, it follows that every SOCP can be expressed as an SDP since for $(x, t) \in \mathbb{R}^{n+1}$ with $t \neq 0$ we have

$$||x|| \leqslant t \;\Leftrightarrow\; x^T x \leqslant t^2 \;\Leftrightarrow\; \begin{pmatrix} tI & x \\ x^T & t \end{pmatrix} \succeq 0.$$

The same equivalences hold for $t = 0$, as in each of the three inequalities this implies $x = 0$.

Thus, we can solve SOCPs using SDP solvers. However, this is a suboptimal approach as there are faster algorithms that are tailored to solve SOCPs [70].

In contrast to LP and SOCP, it is not clear whether we can always find a solution of an SDP in polynomial time. In fact, Ramana [84] gave an example for which every feasible solution has double exponential bit-size, i.e. requires coefficients of size roughly 2^{2^n} where the matrix variable has size $n \times n$, even though the problem looks simple and has only $O(n)$ constraints. It turns out that this example is a rather special case since it is indeed possible to compute a solution of an SDP up to arbitrary accuracy in polynomial time under mild assumptions. This has been shown by Grötschel, Lovász and Schrijver in 1981 [43] using the ellipsoid method. In 2016 de Klerk and Vallentin [31] gave an alternative proof based on an interior point method and the corresponding analysis that is due to Nesterov and Nemirovski [80].

Theorem 3.2.8. *[43],[31] Consider the primal problem in* (3.5)*. Assume that $C, A_1, \ldots, A_m \in \mathbb{Q}^{n \times n}$ and that $b_1, \ldots, b_m \in \mathbb{Q}$. Let*

$$L = \{X \in \mathcal{S}^n \colon \langle A_i, X \rangle = b_i \;\; \forall\, i \in \{1, \ldots, n\}\} \text{ and}$$
$$\mathcal{F} = \{X \in L \colon X \succeq 0\}.$$

Furthermore, assume that there is a matrix $X \in \mathcal{F} \cap \mathbb{Q}^{n \times n}$ and $r, R \in \mathbb{Q}$ such that

$$B_{X_0}(r) \cap L \subseteq \mathcal{F} \subseteq B_{X_0}(R) \cap L,$$

where $B_{X_0}(r) \subseteq \mathcal{S}^n$ is the ball of radius r and center X_0 with respect to the metric that is induced by $\langle \cdot, \cdot \rangle$. Then for every $\varepsilon \in \mathbb{Q}_{>0}$ it is possible to find an X^* with

$$\langle C, X^* \rangle - p^* \leqslant \varepsilon,$$

in polynomial time. The polynomial is in $n, m, \log_2(R/r), \log_2(1/\varepsilon)$ and the bit size of $C, A_1, \ldots, A_m, b_1, \ldots, b_m$ and X_0.

3.2.4 Completely positive programming

The last special case that we pay particular attention to employs the *completely positive cone*:

$$\mathcal{CP}_n := \text{cone}\left\{xx^T : x \in \mathbb{R}^n_{\geqslant 0}\right\}.$$

We say that a matrix is *completely positive* if it lies in this cone. On the first glance, $\mathcal{CP}_n$ looks similar to $\mathcal{S}^n_{\geq 0}$. It is the conic hull of positive semidefinite matrices with rank 1. The difference is that these rank-one matrices are additionally required to be componentwise nonnegative in the definition of $\mathcal{CP}_n$. It is remarkable that, while checking positive semidefiniteness is easy, it is NP-hard to determine whether a given matrix is completely positive [32]. In the following we use the notation $X \geqslant 0$ (respectively $X > 0$) to express that the matrix X is componentwise nonnegative (respectively positive). With this consideration we observe that

$$\mathcal{CP}_n \subseteq \mathcal{S}^n_{\geq 0} \cap \mathbb{R}^{n\times n}_{\geqslant 0}. \tag{3.6}$$

It is known that (3.6) holds with equality if and only if $n \leqslant 4$ [73],[41]. Thus, the interior of $\mathcal{CP}_n$ for $n \leqslant 4$ is given by:

$$\operatorname{int}\mathcal{CP}_n = \operatorname{int}\mathcal{S}^n_{\geq 0} \cap \operatorname{int}\mathbb{R}^{n\times n}_{\geqslant 0} = \{X \in \mathcal{S}^n : X \succ 0,\ X > 0\}.$$

This is noted by Dür and Still [34] who additionally give a description of the interior of $\mathcal{CP}_n$ for the general case.

Theorem 3.2.9. *[34] For every $n \in \mathbb{N}$ we have:*

$$\operatorname{int}\mathcal{CP}_n = \left\{XX^T : X = [X_1|X_2],\ \text{where } X_1 \text{ is regular and } X_1 > 0,\ X_2 \geqslant 0\right\}.$$

Next, we specify the setting to state the corresponding primal-dual pair of conic programs:

$$\begin{aligned} V_1 = W_1 &= \mathbb{R}^m, \\ V_2 = W_2 &= \mathcal{S}^n, \\ K_2 &= \mathcal{CP}_n, \\ (u,v)_1 &= u^T v \ \text{ for all } u, v \in \mathbb{R}^m, \\ (X,Y)_2 &= \langle X, Y\rangle \ \text{ for all } X, Y \in \mathcal{S}^n. \end{aligned}$$

We employ the same notation for a linear map $A\colon \mathcal{S}^n \to \mathbb{R}^m$ as in Chapter 3.2.3. Unlike in the previous cases, $\mathcal{CP}_n$ is not self-dual. Instead, we have

$$\mathcal{CP}^*_n = \mathcal{COP}_n := \left\{X \in \mathcal{S}^n : y^T X y \geqslant 0 \ \ \forall\, y \in \mathbb{R}^n_{\geqslant 0}\right\}.$$

The proof for this is straightforward. The cone $\mathcal{COP}_n$ is called the *copositive cone*. Similarly, we call the matrices contained in $\mathcal{COP}_n$ *copositive*. From the definition it is immediate that

$$\mathcal{S}^n_{\geq 0} + \mathbb{R}^{n\times n}_{\geqslant 0} \subseteq \mathcal{COP}_n.$$

Since $\mathcal{COP}_n$ and $\mathcal{CP}_n$ form a dual pair, Dür and Still [34] state some properties of $\mathcal{COP}_n$ as well, including a description of its interior.

Theorem 3.2.10. *[34] For every $n \in \mathbb{N}$ we have:*

$$\operatorname{int} \mathcal{COP}_n = \left\{ X \in \mathcal{S}^n \colon x^T X x > 0 \ \ \forall\, x \in \mathbb{R}^n_{\geqslant 0} \backslash \{0\} \right\}.$$

As before we plug this setting into (3.3). We get:

$$\begin{array}{lll} p^* = \sup \langle C, X \rangle & & d^* = \inf y^T b \\ \quad \langle A_i, X \rangle = b_i \ \ \forall\, 1 \leqslant i \leqslant m & \text{and} & \quad \sum_{i=1}^{m} y_i A_i - C \in \mathcal{COP}_n \\ \quad X \in \mathcal{CP}_n & & \quad y \in \mathbb{R}^m. \end{array} \tag{3.7}$$

We call this type of conic programming *completely positive programming* (CP). In the literature the term "copositive programming" is also common and refers to the same primal-dual pair of conic programs. We use these terms interchangeably.

The difficulty of CP differs greatly from the difficulty of LP, SOCP and SDP. While LP, SOCP and SDP (under mild assumptions) can be solved in polynomial time, this is not the case for CP. Solving CP is an NP-hard problem because it is NP-hard to determine whether a given matrix is contained in CP, as mentioned earlier. At the same time, CP is a meaningful type of optimization problems. Burer [25] showed that every quadratic optimization problem with binary and/or continuous variables can be expressed as a CP.

One way to address this difficulty is to search for good upper and lower bounds instead of solving CP exactly. Since the primal problem in (3.7) is a maximization problem, any feasible solution yields a lower bound on the optimum. Finding upper bounds is a little less straightforward. One possibility is to make use of (3.6) and the fact that SDP is usually easier to solve than CP:

$$\begin{array}{lcl} \sup \langle C, X \rangle & \leqslant & \sup \langle C, X \rangle \\ \quad \langle A_i, X \rangle = b_i \ \ \forall\, 1 \leqslant i \leqslant m & & \quad \langle A_i, X \rangle = b_i \ \ \forall\, 1 \leqslant i \leqslant m \\ \quad X \in \mathcal{CP}_n & & \quad X \succeq 0,\ X \geqslant 0. \end{array} \tag{3.8}$$

Note that we can bring the right-hand side of (3.8) into the same form as in (3.5) by introducing slack variables.

Chapter Four

An algorithm to compute the EHZ capacity of Lagrangian products

The goal of this chapter is to provide an algorithm that computes $c_{\mathrm{EHZ}}(C)$ (see Definition 2.3.5) or a corresponding upper bound, where $C \subseteq \mathbb{R}^{2n}$ is a Lagrangian product. This means, there are convex sets $K, T \subseteq \mathbb{R}^n$ such that $C = K \times T$. To do so, we examine the concept of Minkowski billiards. The connection between Minkowski billiards and the EHZ capacity was first studied by Artstein-Avidan and Ostrover in 2012 [11]. Additionally, we state some properties of Minkowski billiards that allow us to formulate an algorithm to compute a closed Minkowski billiard trajectory with minimal length. Consequently, this yields an algorithm to compute the EHZ capacity of $K \times T$. We develop our algorithm first in the Euclidean setting, where K is a polytope and T is the Euclidean unit ball $B_1(0)$. Afterwards, we discuss the algorithm in the Minkowski setting where both K and T are polytopes.

This chapter is based on the papers [65] and [66]. The latter paper is currently in preparation. The theoretical aspects of this chapter, which are presented in Chapters 4.1 and 4.2, are mainly attributed to Daniel Rudolf. Therefore, we omit the proofs here and refer to [65] and [66].

4.1 From the EHZ capacity to closed Minkowski billiard trajectories

To understand the term billiard as it is used in mathematics we can think of an actual game of billiards. If we play a ball on the billiard table (without doing trick shots), then the ball moves straight and only changes directions when bouncing of a wall. In other words, the route of the ball is a polygonal line. Here, a *polygonal line* is a curve $\gamma\colon [0, \infty) \to \mathbb{R}^n$ such that there is $t_i \in \mathbb{R}$ for every $i \in \mathbb{N}$ with:

(*i*) $t_0 = 0$ and $t_{i+1} > t_i$ for every $i \in \mathbb{N}$.

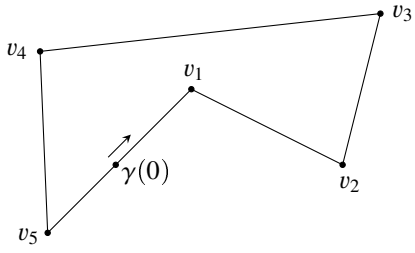
(*ii*) $\gamma|_{[t_i, t_{i+1}]}$ is linear for every $i \in \mathbb{N}$.

(*iii*) For every $i \in \mathbb{N}$ the points $\gamma(t_i), \gamma(t_{i+1})$ and $\gamma(t_{i+2})$ are not contained in a common straight line.

We call the points $\gamma(t_0), \gamma(t_1), \ldots$ the vertices of γ. If γ is periodic, say with period $a \in \mathbb{R}_{>0}$, then we say that γ is closed. In this case, if we let

$$\{t_{i_1}, \ldots, t_{i_m}\} = \{t_i \colon i \in \mathbb{N}\} \cap [a, 2a],$$

then we refer to $\gamma(t_{i_1}), \ldots, \gamma(t_{i_m})$ as the vertices of γ, because these vertices repeat themselves. As an exception, if γ is a closed curve that moves linearly from some point $u \in \mathbb{R}^n$ to some point $v \in \mathbb{R}^n$, then back from v to u, then from u to v again and so forth, we say that γ is a closed polygonal line with the vertices u and v, even though property (*iii*) is violated. If we refer to a closed polygonal line γ as a subset of $\mathbb{R}^n$, then we mean the image of γ.



$$\begin{aligned} v_1 &= \gamma(t_1) = \gamma(t_6) = \ldots \\ v_2 &= \gamma(t_2) = \gamma(t_7) = \ldots \\ v_3 &= \gamma(t_3) = \gamma(t_8) = \ldots \\ v_4 &= \gamma(t_4) = \gamma(t_9) = \ldots \\ v_5 &= \gamma(t_5) = \gamma(t_{10}) = \ldots \end{aligned}$$

Figure 4.1: A closed polygonal line. Its period a lies between t_5 and t_6. On the interval $[a, 2a]$ we can find the points t_6, t_7, t_8, t_9 and t_{10}. Thus, γ has 5 vertices.

If we think of a game of billiards again, the vertices of the route of the billiard ball are the points at which the ball hits a wall. At these points the route of the ball follows the Euclidean billiard reflection rule: The angle of reflection equals the angle of incidence. The upcoming definition generalizes this concept in a sense that the reflection rule is governed by a convex set T. We make a few remarks before we state the definition formally.

We recall from Chapter 2.3.1 that for a set $T \subseteq \mathbb{R}^n$ the *Minkowski functional* $\mu_T \colon \mathbb{R}^n \to \mathbb{R}$ is defined by:

$$\mu_T(y) = \inf\{t \geqslant 0 \colon y \in tT\} \quad \text{for all } y \in \mathbb{R}^n,$$

where we can replace the infimum by a minimum if T is compact. For two points $b, d \in \mathbb{R}^n$ we let $[b, d]$ denote the line segment with end points b and d. Furthermore, whenever we consider the vertices $v_1, \ldots, v_m$ of a closed polygonal line γ in this chapter, these vertices are ordered such that $[v_i, v_{i+1}]$ is contained in the image of γ. Note that this ordering is not unique because if the ordering $v_1, \ldots, v_m$ satisfies this property, then so do

$$\begin{aligned} &v_{(1+k) \bmod m},\ v_{(2+k) \bmod m},\ \ldots,\ v_{(m+k) \bmod m} \quad \text{and} \\ &v_{(m+k) \bmod m},\ \ldots,\ v_{(2+k) \bmod m},\ v_{(1+k) \bmod m} \end{aligned}$$

for every $k \in \{1, \ldots, m\}$.

Definition 4.1.1. *Let $K, T \subseteq \mathbb{R}^n$ be convex bodies and let $m \geqslant 2$.*

(i) A closed polygonal line with vertices $q_1, \dots, q_m \in \partial K$ is called a closed (K,T)-Minkowski billiard trajectory (MBT) if for every $j \in \{1, \dots, m\}$ there is a supporting hyperplane H_j of K with $q_j \in H_j$ such that:

$$q_j = \operatorname{argmin} \{\mu_{T^\circ}(\bar{q}_j - q_{j-1}) + \mu_{T^\circ}(q_{j+1} - \bar{q}_j) \colon \bar{q}_j \in H_j\},$$

where $q_0 := q_m$ and $q_{m+1} := q_1$.

(ii) The vertices of a closed (K,T)-MBT are called bouncing points.

(iii) A closed (K,T)-MBT is called regular if all its bouncing points $q_1, \dots, q_m$ are smooth boundary points, i.e. $N_K(q_j)$ is one-dimensional for every $j \in \{1, \dots, m\}$.

Recall that the normal cone is given by

$$N_K(y) = \{u \in \mathbb{R}^n \colon u^T(z - y) \leqslant 0 \text{ for all } z \in K\},$$

as defined in Chapter 2.3.1. We proceed with some remarks regarding the notation. We write "closed MBT" instead of "closed (K,T)-MBT" if K and T are clear from context. Furthermore, we encode a closed polygonal line with vertices $q_1, \dots, q_m$ by $(q_1, \dots, q_m)$ throughout this chapter. We let $q_j := q_{j \bmod m}$ for every $j \in \mathbb{Z}$. If we have $T = B_1(0)$, it is customary to call a closed Minkowski billiard trajectory a *closed Euclidean billiard trajectory* (EBT) instead. In this case the Minkowski functional is the Euclidean norm and one verifies that for the minimum of

$$||\bar{q}_j - q_{j-1}|| + ||q_{j+1} - \bar{q}_j||,$$

the angle of reflection equals the angle of incidence. Therefore, we recover the Euclidean billiard reflection rule.

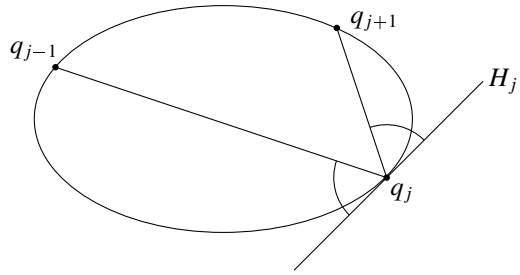


Figure 4.2: Three consecutive bouncing points of a closed EBT. The two indicated angles (i.e. angle of reflection and angle of incidence at q_j) are equal. The boundary of the billiard table K is an ellipse.

Billiards have broad application and they are studied actively. For an overview on billiards we refer to [45], [61] and [93]. In particular, Euclidean billiards have been studied intensively in dimension two. On the other hand, not much is known about Euclidean

billiards in higher dimension. Even less results are known for the more general Minkowski case where T is an arbitrary convex set. To our knowledge, the only publications that treat Minkowski billiards intensively are [2], [3], [4], [5], [6], [9], [11], [20], [46] and [83].

A particularly important, known fact about Minkowski billiards is that we can characterize them much easier than Definition 4.1.1 suggests.

Theorem 4.1.2. *[66] Let $K \subseteq \mathbb{R}^n$ be a convex body and let $T \subseteq \mathbb{R}^n$ be a strictly convex body. Then a closed polygonal line $(q_1, \dots, q_m)$ with $q_j \in \partial K$ for every $j \in \{1, \dots, m\}$ is a closed MBT if and only if there are points $p_1, \dots, p_m$ on ∂T such that*

$$\left.\begin{cases} q_{j+1} - q_j \in N_T(p_j), \\ p_{j+1} - p_j \in -N_K(q_{j+1}) \end{cases}\right\} \quad \text{for all } j \in \{1, \dots, m\}, \tag{4.1}$$

where $q_{m+1} := q_1$ and $p_{m+1} := p_1$.

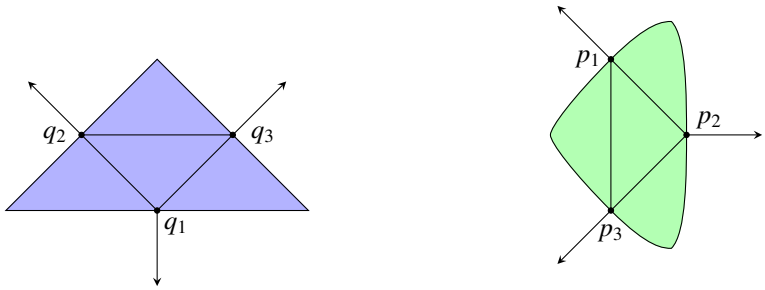


Figure 4.3: A closed MBT with vertices q_1, q_2, q_3 and corresponding points p_1, p_2, p_3. The convex body K is depicted in blue and the strictly convex body T is depicted in green. Also shown are the normal cones at q_1, q_2, q_3 and p_1, p_2, p_3. One can see that (4.1) is satisfied.

Theorem 4.1.2 shows that with each closed MBT $q = (q_1, \dots, q_m)$ we can associate a (not necessarily unique) tupel $p = (p_1, \dots, p_m)$ of boundary points of T. We call p a *dual* closed MBT of q. Note that we do not require that consecutive points in $p_1, \dots, p_m$ are distinct. In fact, Figure 4.4 depicts an example in which there are two consecutive points in p that are equal. Thus, strictly speaking, p does not encode a closed polygonal line unless we allow that line segments between two vertices can have length 0. This ambiguity has not much influence on the remainder of this thesis. We simply keep in mind that consecutive vertices of closed MBTs are distinct while consecutive vertices of dual closed MBTs may not be distinct.

For a closed polygonal line $(q_1, \dots, q_m)$ we call

$$\ell_T((q_1, \dots, q_m)) = \sum_{j=1}^{m} \mu_{T^\circ}(q_{j+1} - q_j)$$

with $q_{m+1} := q_1$ the *ℓ_T-length* of q. Another discovery about Minkowski billiards, which is highly relevant for this thesis, is the fact that the minimal ℓ_T-length of a closed MBT is exactly the minimal action of a generalized closed characteristic (see Definition 2.3.5).

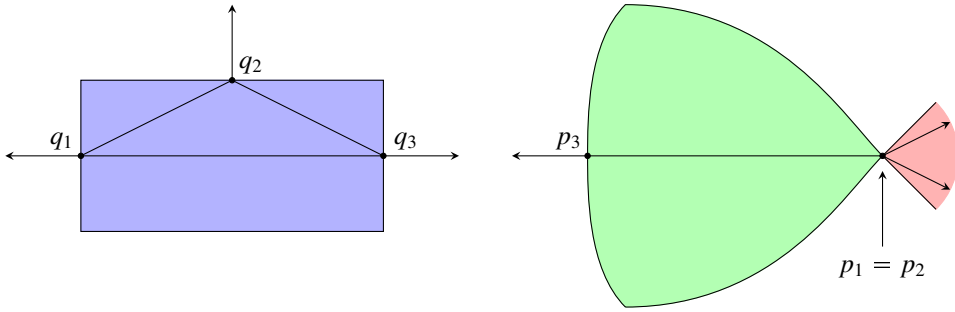


Figure 4.4: A closed MBT $q = (q_1, q_2, q_3)$ and a dual closed MBT (p_1, p_2, p_3) of q. Not all consecutive vertices of the dual closed MBT are distinct. More precisely, $p_1 = p_2$. As in Figure 4.3 K is depicted in blue and T is depicted in green. The red area depicts the normal cone at p_1 or p_2.

Theorem 4.1.3. *[66] Let $K \subseteq \mathbb{R}^n$ be a convex body and let $T \subseteq \mathbb{R}^n$ be a strictly convex body.*

(i) Let x be a generalized closed characteristic on $\partial(K \times T)$ with minimal action, i.e. $c_{\mathrm{EHZ}}(K \times T) = \mathbb{A}(x)$. Then there is a generalized closed characteristic $\widetilde{x} = (\widetilde{x}_q, \widetilde{x}_p)$ on $\partial(K \times T)$ such that:

- *$\widetilde{x}_q$ is a closed (K, T)-MBT with minimal ℓ_T-length,*
- *$\widetilde{x}_p$ is a dual closed (K, T)-MBT of $\widetilde{x}_q$,*
- *$\mathbb{A}(x) = \mathbb{A}(\widetilde{x}) = \ell_T(\widetilde{x}_q)$.*

(ii) Let q be a closed MBT with minimal ℓ_T-length. Then there is a generalized closed characteristic x on $\partial(K \times T)$ with minimal action and

$$\ell_T(q) = \mathbb{A}(x).$$

Theorem 4.1.3 makes apparent that it is beneficial to compute the ℓ_T-length of some closed MBTs if we want to study the EHZ capacity. Moreover, if we are given the vertices of a closed MBT q as well as the vertices of a dual closed MBT, then the computation of the ℓ_T-length of q is particularly simple.

Lemma 4.1.4. *[66] Let $T \subseteq \mathbb{R}^n$ be a convex body. Then for every $z \in \mathbb{R}^n$ and every $y \in \partial T$ we have:*

$$\mu_{T^\circ}(z) = z^T y \iff z \in N_T(y).$$

If $q = (q_1, \ldots, q_m)$ is a closed MBT and if p is a dual closed MBT of q, then we have $q_{j+1} - q_j \in N_T(p_j)$. Consequently, if we denote $q_{m+1} := q_1$, the ℓ_T-length of q is

$$\ell_T(q) = \sum_{j=1}^{m} \mu_{T^\circ}(q_{j+1} - q_j) = \sum_{j=1}^{m} (q_{j+1} - q_j)^T p_j. \tag{4.2}$$

Now, we can outline the strategy that we pursue in this chapter to calculate $c_{\mathrm{EHZ}}(K \times T)$. From Theorem 4.1.3 we know that finding a generalized closed characteristic with minimal action (and consequently finding $c_{\mathrm{EHZ}}(K \times T)$) is equivalent to finding a closed MBT with minimal ℓ_T-length. Additionally, Theorem 4.1.2 provides us with a way to find such a closed MBT. We search for points $q_1, \dots, q_m \in \partial K$ and $p_1, \dots, p_m \in \partial T$ that satisfy (4.1). Alkoumi and Schlenk [6] also followed this approach in the Euclidean case for two-dimensional sets and suggested the application of this idea to the more general Minkowski setting as an open problem.

4.2 Properties of closed Minkowski billiard trajectories

The goal is now to use the strategy that we sketch at the end of Chapter 4.1 to formulate an algorithm that computes $c_{\mathrm{EHZ}}(K \times T)$. In the process we encounter some problems that we need to solve first. The most obvious problem is that we successively search for m points on the boundary of K and m points on the boundary of T for every $m \geqslant 2$. This is difficult without a threshold for m at which we can stop. In [6] this problem is resolved using a result from [18] that states that a closed EBT with minimal ℓ_T-length has only two or three bouncing points. The following theorem is a generalization of this result to the Minkowski setting.

Theorem 4.2.1. *[66] Let $K \subseteq \mathbb{R}^n$ be a convex body and let $T \subseteq \mathbb{R}^n$ be a strictly convex body with smooth boundary. Furthermore, let $(q_1, \dots, q_m)$ be a closed MBT with minimal ℓ_T-length. Following Theorem 4.1.2 we take $p_j \in \partial T$, $u_j \in N_K(q_j)$ with $||u_j|| = 1$ and $\mu_j \in \mathbb{R}_{\geqslant 0}$ such that*

$$p_{j+1} - p_j = -\mu_{j+1} u_{j+1},$$

for every $j \in \{1, \dots, m\}$, where $p_{m+1} := p_1$, $u_{m+1} := u_1$ and $\mu_{m+1} := \mu_1$. Then the cone

$$\mathcal{U} := \operatorname{cone}\{u_1, \dots, u_m\}$$

is a linear space with dimension $m - 1$. Moreover, we have

$$\dim(N_K(q_j) \cap \mathcal{U}) = 1,$$

for every $j \in \{1, \dots, m\}$.

Theorem 4.2.1 is useful for the implementation of our algorithm for three reasons. First, we have

$$m - 1 = \dim \mathcal{U} \leqslant n,$$

which means that in our algorithm we only need to consider $m \in \{2, \dots n + 1\}$ rather than every $m \geqslant 2$.

Second, if we consider $m = n + 1$, then Theorem 4.2.1 states that the normal cone $N_K(q_j)$ is one-dimensional for every $j \in \{1, \dots, m\}$. Hence, if we search for a closed MBT with minimal ℓ_T-length that has $n + 1$ bouncing points, then it is sufficient to look for a *regular* closed MBT.

Third, Theorem 4.2.1 gives us a way to rule out certain choices of $q_1, \dots, q_m$ in advance. Let us assume we are given $q_1, \dots, q_m$ that are smooth boundary points of K. This means the normal cone $N_K(q_j)$ is one-dimensional for each $j \in \{1, \dots, m\}$. Therefore, $u_1, \dots, u_m$ are uniquely determined and we can check whether the cone spanned by them is a linear space with dimension $m - 1$. If not, we know that $(q_1, \dots, q_m)$ is not a closed MBT and there is no need to look for $p_1, \dots, p_m$.

We have another theorem that allows for a similar observation. It makes a statement on the dimension of the convex hull of the bouncing points $q_1, \dots, q_m$. Via Theorem 4.1.2 this turns out to be useful in the search of a dual closed MBT $(p_1, \dots, p_m)$.

Theorem 4.2.2. *[66] Let $K, T \subseteq \mathbb{R}^n$ be strictly convex bodies with smooth boundaries and let $q = (q_1, \dots, q_m)$ be a closed MBT with minimal ℓ_T-length. Then we have*

$$\dim(\operatorname{conv}\{q_1, \dots, q_m\}) = m - 1.$$

Later in Chapter 4.3.1 we see how Theorem 4.2.1 and Theorem 4.2.2 can be exploited in the implementation of our algorithm.

Next, we discuss an alternative formulation for closed MBTs with minimal ℓ_T-length. For this we let

$$F(K) := \{M \subseteq \mathbb{R}^n \colon \nexists\, t \in \mathbb{R}^n,\ M + t \subseteq \operatorname{int} K\},$$
$$F_{n+1}(K) := \{M \subseteq \mathbb{R}^n \colon |M| \leqslant n + 1,\ M \in F(K)\}.$$

In other words, $F_{n+1}(K)$ consists of sets with at most $n + 1$ points that we cannot shift into the interior of K. We have the following theorem that establishes a one-to-one correspondence between closed MBTs with minimal ℓ_T-length and members of $F_{n+1}(K)$ with minimal ℓ_T-length.

Theorem 4.2.3. *[66] Let $K \subseteq \mathbb{R}^n$ be a convex body and let $T \subseteq \mathbb{R}^n$ be a strictly convex body. Furthermore, let $q = (q_1, \dots, q_m)$ be a closed polygonal line. Then q is a closed MBT with minimal ℓ_T-length if and only if it is a member of*

$$\{q' = (q'_1, \dots, q'_k) \colon k \geqslant 2,\ \{q'_1, \dots, q'_k\} \in F_{n+1}(K)\}$$

with minimal ℓ_T-length.

We use Theorem 4.2.3 in Chapter 4.3.2 to justify the application of our algorithm to instances where T is a polytope and thus neither strictly convex nor with smooth boundary. More precisely, we use Theorem 4.2.3 together with a continuity result to show that it is sufficient to find a member of $F_{n+1}(K)$ with minimal ℓ_T-length if we want to compute $c_{\mathrm{EHZ}}(K \times T)$, even if T is not strictly convex. We give the continuity result in the following theorem. Every convergence that comes up in this statement is with respect to the Hausdorff metric d_H as defined in Chapter 2.3.

Theorem 4.2.4. *[66]*

(i) Let $T \subseteq \mathbb{R}^n$ be a strictly convex body and let $(K_i)_{i \in \mathbb{N}}$ be a sequence of convex bodies in $\mathbb{R}^n$ that converges to some convex body $K \subseteq \mathbb{R}^n$. Then there is a strictly increasing sequence $(i_j)_{j \in \mathbb{N}}$ of indices such that the sequence $(q^{i_j})_{j \in \mathbb{N}}$ of closed (K_{i_j}, T)-MBTs with minimal ℓ_T-length converges to a closed (K, T)-MBT with minimal ℓ_T-length.

(ii) Let $K \subseteq \mathbb{R}^n$ be a convex body and let $(T_i)_{i\in\mathbb{N}}$ be a sequence of strictly convex bodies in $\mathbb{R}^n$ that converges to some convex body $T \subseteq \mathbb{R}^n$. Then there is a strictly increasing sequence $(i_j)_{j\in\mathbb{N}}$ of indices such that the sequence $(q^{i_j})_{j\in\mathbb{N}}$ of closed (K, T_{i_j})-MBTs with minimal $\ell_{T_{i_j}}$-length converges to a closed polygonal line $q = (q_1, \dots, q_m)$ with $\{q_1, \dots, q_m\} \in F(K)$ and

$$\ell_T(q) = \min\left\{\ell_T(q') \colon q' = (q'_1, \dots q'_k),\ \{q'_1, \dots, q'_k\} \in F(K)\right\}.$$

Furthermore, q is a closed (K, T)-MBT.

4.3 Formulation and implementation of the algorithm

Next, we turn our attention to the implementation of the strategy that we sketch at the end of Chapter 4.1. We start with the Euclidean case and develop an algorithm to find a regular closed EBT that has minimal ℓ_T-length among all regular closed EBTs. According to Theorem 4.1.3 this gives us an upper bound on the EHZ capacity of $K \times B_1(0)$.

Afterwards, we focus on the Minkowski case and formulate an algorithm to find a closed MBT with minimal ℓ_T-length if both K and T are two-dimensional polytopes, i.e. polygons. Thus, this gives us an algorithm that computes the EHZ capacity of $K \times T$ for polygons K and T.

4.3.1 The Euclidean setting

In the following, we let $K \subseteq \mathbb{R}^n$ be a full-dimensional polytope. Since we consider the case $T = B_1(0)$, the ℓ_T-length is the usual Euclidean distance. Thus, we simply say length instead of ℓ_T-length while we examine the Euclidean setting. First, we recall the strategy that we sketch at the end of Chapter 4.1 and provide a few more details.

Algorithm 1 Shortest closed EBT

1: **for** $m \in \{2, \dots, n+1\}$ **do**
2: **for** every choice of m pairwise different facets $F_1, \dots, F_m$ of K **do**
3: For every $j \in \{1, \dots, m\}$ let u_j be a unit vector in $N_K(q_{j+1})$ for some point q_{j+1} in the relative interior of F_{j+1}, where $q_{m+1} := q_1$ and $F_{m+1} := F_1$. Construct a closed polygonal line γ by successively moving in the directions $-u_1, \dots, -u_m$. Translate and scale γ with a positive factor such that its vertices lie on the unit sphere S^{n-1}. Let $n_1, \dots, n_m \in S^{n-1}$ be these vertices, i.e.:

$$n_{j+1} - n_j = -\mu_j u_j$$

with appropriate $\mu_j \in \mathbb{R}_{>0}$ for every $j \in \{1, \dots, m\}$, where $n_{m+1} := n_1$.
4: Construct a closed polygonal line ξ by successively moving in the directions $n_1, \dots, n_m$. Translate and scale ξ with a positive factor to get a closed polygonal line $q = (q_1, \dots, q_m)$ such that $q_j \in F_j$ and

$$q_{j+1} - q_j = \lambda_j n_j$$

with appropriate $\lambda_j \in \mathbb{R}_{>0}$ for every $j \in \{1, \dots, m\}$, where $q_{m+1} := q_1$.

5: Calculate the length of the closed polygonal line q and store it, if it is the smallest among all such closed polygonal lines found so far.
6: **end for**
7: **end for**

Note that the unit vector u_j that appears in Algorithm 1 is unique for every $j \in \{1, \dots, m\}$ since $N_K(q_{j+1})$ is one-dimensional and that u_j does not depend on the choice of q_{j+1}. Furthermore, we point out that it is not always possible to construct the closed polygonal lines γ and ξ or to translate and scale them in the required way. In this case we reject the current iteration and proceed with the next choice of facets $F_1, \dots, F_m$.

Theorem 4.1.2 ensures that every closed polygonal line found by Algorithm 1 is indeed a closed EBT. More precisely, if $(q_1, \dots, q_m)$ is such a closed polygonal line, then it fulfils (4.1) by construction. We will now examine this algorithm in more detail.

First, we note that we require $\mu_j, \lambda_j > 0$ here, while Theorem 4.1.2 only suggests $\mu_j, \lambda_j \geqslant 0$. The reason for this is that, on the one hand, the bouncing points $q_1, \dots, q_m$ should be pairwise distinct. Thus, we can consider positive instead of nonnegative λ_j. On the other hand, if $\mu_i = 0$ for some $i \in \{1, \dots, m\}$, then we have $n_{i+1} = n_i$. Thus, $q_{i+2} - q_{i+1}$ is a multiple of $q_{i+1} - q_i$, where $q_{m+1} := q_1$ and $q_{m+2} := q_2$. In other words, we find that q_i, q_{i+1} and q_{i+2} lie on a straight line. Therefore, we can remove q_{i+1} and n_{i+1} without changing the length of $(q_1, \dots, q_m)$. One can check that after removing q_{i+1} and n_{i+1} the remaining points still fulfil (4.1). So, the algorithm already found a closed EBT with less than m bouncing points that is at least as short as $(q_1, \dots, q_m)$.

We proceed to discuss the third step in Algorithm 1. For this, we let $F_1, \dots, F_m$ and $u_1, \dots, u_m$ be as described above in the algorithm. Let U be the $(n \times m)$-matrix containing $u_1, \dots, u_m$ as columns. If for every $j \in \{1, \dots, m\}$ there is a smooth boundary point q_j of T with $q_j \in F_j$ and $n_j \in S^{n-1}$ such that (4.1) is fulfilled, then there are $\mu_1, \dots, \mu_m > 0$ such that $n_{j+1} - n_j = -\mu_j u_j$ with $n_{m+1} := n_1$. Thus, because of

$$0 = \sum_{j=1}^{m} (n_{j+1} - n_j) = -\sum_{j=1}^{m} \mu_j u_j \,,$$

we have that $u_1, \dots, u_m$ are linearly dependent. Hence, $\operatorname{rk}(U) \leqslant m - 1$. On the other hand, if we assume that the regular closed EBT $(q_1, \dots, q_m)$ has minimal length, then by Theorem 4.2.1 the convex cone spanned by $u_1, \dots, u_m$ is an $(m-1)$-dimensional linear space. In particular, there are $(m-1)$ linearly independent vectors in $\{u_1, \dots, u_m\}$, so $\operatorname{rk}(U) \geqslant m - 1$. Together, we have that $\operatorname{rk}(U) = m - 1$ is necessary if we search for a regular closed EBT with minimal length. In this way, some choices of $F_1, \dots, F_m$ can be discarded immediately. More precisely, after we choose facets $F_1, \dots, F_m$ we compute $\operatorname{rk}(U)$. If this rank is not equal to $m - 1$, then we can directly proceed with the next choice of facets. We note that $\operatorname{rk}(U) = m - 1$ also implies that

$$-\sum_{j=1}^{m} \mu_j u_j = 0$$

has up to scaling a unique solution $\mu_1, \dots, \mu_m$. Consequently, there is up to homothetic transformations (i.e. up to scaling and translating) only one closed polygonal line that we can get by successively moving in the directions $-u_1 \dots, -u_m$.

To find suitable $n_1, \dots, n_m \in S^{n-1}$ we let $\gamma = (\gamma_1, \dots, \gamma_m)$ be this closed polygonal line. The task is now to translate and scale γ with a positive factor such that the vertices of γ lie on S^{n-1}. We take $n_1, \dots, n_m$ as these vertices. Note that in the fourth step of the algorithm it is required to form another closed polygonal line by successively moving in the directions $n_1, \dots, n_m$, i.e.:

$$\exists \lambda_1, \dots, \lambda_m > 0 \colon \sum_{j=1}^{m} \lambda_j n_j = 0. \tag{4.3}$$

If this property is satisfied, one says that $n_1, \dots, n_m$ form a *totally cyclic vector configuration*. Following [105] we can find an equivalent characterization by using Farkas' lemma. More precisely, (4.3) holds if and only if for every vector $v \in \mathbb{R}^n$ one of the following two conditions is satisfied:

(a) $n_j^T v < 0$ for some $j \in \{1, \dots, m\}$,

(b) $n_j^T v = 0$ for all $j \in \{1, \dots, m\}$.

Hence, if $n_1, \dots, n_m$ are a totally cyclic vector configuration, it follows that

$$\forall\, v \in \mathbb{R}^n \;\exists\; j \in \{1, \dots, m\} \,:\, v^T n_j \leqslant 0. \tag{4.4}$$

This property is less restrictive than (4.3), which is illustrated in Figure 4.5, but it is still sufficient for the upcoming arguments. In particular, we would like to scale and translate γ such that its vertices $n_1, \dots, n_m \in S^{n-1}$ satisfy (4.4). If this is not possible, then we are not able to construct the closed polygonal line ξ in the fourth step of Algorithm 1 and consequently we can reject the current iteration.

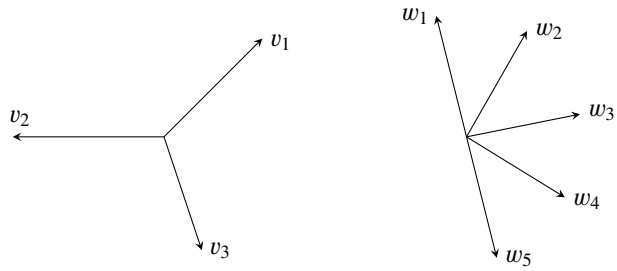


Figure 4.5: On the left, the points v_1, v_2, v_3 form a totally cyclic vector configuration. On the right, the points w_1, w_2, w_3, w_4, w_5 do not form a totally cyclic vector configuration but they still satisfy (4.4).

While there might be multiple possibilities to scale and translate γ such that its vertices lie on S^{n-1}, there is at most one possibility such that the vertices fulfil (4.4).

Lemma 4.3.1. *Let* $\gamma = (\gamma_1, \ldots, \gamma_m) \in \mathbb{R}^n \times \ldots \times \mathbb{R}^n$,

$$S_\gamma = \{(\mu, t) \colon \mu \in \mathbb{R}_{>0},\, t \in \mathbb{R}^n,\, ||\mu\gamma_j + t|| = 1 \text{ for all } j \in \{1, \ldots, m\}\}$$

and $(\mu^*, t^*) \in S_\gamma$. *If* $n_1, \ldots, n_m \in S^{n-1}$ *satisfy* (4.4), *where* $n_j = \mu^*\gamma_j + t^*$ *for every* j, *then:*

$$\mu^* = \max\{\mu \colon (\mu, t) \in S_\gamma \text{ for some } t \in \mathbb{R}^n\}.$$

Proof. First, we observe that the maximum in the statement is indeed a maximum. To see this, we let $((\mu_k, t_k))_{k\in\mathbb{N}}$ be a convergent sequence in $\mathbb{R} \times \mathbb{R}^n$ with $(\mu_k, t_k) \in S_\gamma$ for every $k \in \mathbb{N}$ such that

$$\hat{\mu} := \lim_{k\to\infty} \mu_k = \sup\{\mu \colon (\mu, t) \in S_\gamma \text{ for some } t \in \mathbb{R}^n\}. \tag{4.5}$$

Since $||\mu\gamma_j + t||$ is continuous in μ and t for every $j \in \{1, \ldots, m\}$, the limit $(\hat{\mu}, \hat{t})$ of this sequence is either contained in S_γ as well or satisfies $\hat{\mu} = 0$. But due to the existence of $(\mu^*, t^*) \in S_\gamma$ we know that S_γ is not empty and in particular $\hat{\mu} > 0$. Hence, the supremum in (4.5) is infact a maximum.

Next, we show that the existence of $(\tilde{\mu}, \tilde{t}) \in S_\gamma$ with $\tilde{\mu} > \mu^*$ yields a contradiction. For $j \in \{1, \ldots, m\}$ we have

$$\tilde{\mu}\gamma_j + \tilde{t} = \tilde{\mu}\frac{n_j - t^*}{\mu^*} + \tilde{t} = \mu' n_j + t',$$

where $\mu' = \tilde{\mu}/\mu^* > 1$ and $t' = \tilde{t} - (\tilde{\mu}/\mu^*)t^*$. Because of $(\mu^*, t^*), (\tilde{\mu}, \tilde{t}) \in S_\gamma$ we get

$$\begin{aligned} 1 = ||\tilde{\mu}\gamma_j + \tilde{t}||^2 = ||\mu' n_j + t'||^2 &= \mu'^2||n_j||^2 + 2\mu' n_j^T t' + ||t'||^2 \\ &= \mu'^2 + 2\mu' n_j^T t' + ||t'||^2. \end{aligned}$$

Since $n_1, \ldots, n_m$ satisfy (4.4), there is some $i \in \{1, \ldots, m\}$ such that $n_i^T(-t') \leqslant 0$. Hence,

$$0 = \mu'^2 + 2\mu' n_i^T t' + ||t'||^2 - 1 \geqslant \mu'^2 + ||t'||^2 - 1 > ||t'||^2 \geqslant 0,$$

which is a contradiction. □

Lemma 4.3.1 implies that if the vertices of $\mu_1\gamma + t_1$ and $\mu_2\gamma + t_2$ with $(\mu_1, t_1), (\mu_2, t_2) \in S_\gamma$ form totally cyclic vector configurations (and hence satisfy (4.4)), then they are scaled by the same factor, i.e. $\mu_1 = \mu_2$. However, we also need $t_1 = t_2$ to make sure that there is only one suitable way to scale and translate γ. The next Lemma shows that this is indeed the case.

Lemma 4.3.2. *Let* $\gamma = (\gamma_1, \ldots, \gamma_m) \in \mathbb{R}^n \times \ldots \times \mathbb{R}^n$, *let* S_γ *as in Lemma 4.3.1 and let* $(\mu_1, t_1), (\mu_2, t_2) \in S_\gamma$. *Furthermore, let* $n_1, \ldots, n_m, n'_1, \ldots, n'_m \in S^{n-1}$ *be defined by*

$$n_j = \mu_1\gamma_j + t_1 \text{ and } n'_j = \mu_2\gamma_j + t_2$$

for every $j \in \{1, \ldots, m\}$. *If* $n_1, \ldots, n_m$ *satisfy* (4.4), *then* $(\mu_1, t_1) = (\mu_2, t_2)$.

Proof. From Lemma 4.3.1 we have $\mu_1 = \mu_2$ and therefore for every $j \in \{1, \dots, m\}$:

$$n_j' = \mu_1 \gamma_j + t_2 = \mu_1 \frac{n_j - t_1}{\mu_1} + t_2 = n_j - t_1 + t_2.$$

Similar to the calculation in the proof of Lemma 4.3.1 we get

$$1 = ||n_j'|| = ||n_j - t_1 + t_2|| = 1 + 2n_j^T(t_2 - t_1) + ||t_2 - t_1||^2.$$

Hence, the term $n_j^T(t_2 - t_1)$ does not depend on j. Because of (4.4) one can find $i \in \{1, \dots, m\}$ such that

$$n_i^T(t_2 - t_1) \leqslant 0$$

and similarly there is $k \in \{1, \dots, m\}$ with

$$n_k^T(t_1 - t_2) \leqslant 0 \iff n_k^T(t_2 - t_1) \geqslant 0.$$

Since these terms do not depend on the indices i and k, we get $n_j^T(t_2 - t_1) = 0$ for every $j \in \{1, \dots, m\}$. Furthermore, we get

$$1 = 1 + ||t_2 - t_1||^2,$$

and so $t_2 = t_1$. □

Assume we find some $n_1, \dots, n_m \in S^{n-1}$ by scaling and translating the closed polygonal line γ as described above. Remember that by construction of γ we have $n_{j+1} - n_j = -\mu_j u_j$ with some $\mu_j > 0$ for every $j \in \{1, \dots, m\}$, where $n_{m+1} := n_1$. This implies:

$$\begin{aligned} 1 = ||n_{j+1}||^2 &= ||n_j - \mu_j u_j||^2 = ||n_j|| - 2\mu_j n_j^T u_j + \mu_j^2 ||u_j||^2 \\ &= 1 - 2\mu_j n_j^T u_j + \mu_j^2 \\ \implies 0 &= \mu_j(\mu_j - 2n_j^T u_j). \end{aligned}$$

Because μ_j is positive, we have $\mu_j = 2n_j^T u_j$ and

$$n_{j+1} = n_j - \mu_j u_j = n_j - 2(n_j^T u_j)u_j$$

for every $j \in \{1, \dots, m\}$. Therefore, if one of the vectors $n_1, \dots, n_m$ is known, then the remaining ones can be calculated recursively. In the following, we choose to search for n_1. This search can be carried out by an SOCP (see Chapter 3.2.2). Before we derive this SOCP we need the following identity:

$$\begin{aligned} n_j &= n_{j-1} - 2(n_{j-1}^T u_{j-1})u_{j-1} = (I - 2u_{j-1}u_{j-1}^T)n_{j-1} \\ &= (I - 2u_{j-1}u_{j-1}^T)(I - 2u_{j-2}u_{j-2}^T)n_{j-2} = \dots = \left(\prod_{i=1}^{j-1} I - 2u_i u_i^T\right) n_1. \end{aligned}$$

This identity holds for every $j \in \{1, \dots, m\}$. Two types of constraints are necessary for the SOCP. First, we rewrite the condition $n_1 = n_{m+1}$ with $n_{m+1} := n_m - 2(n_m^T u_m)u_m$:

$$n_1 = n_m - 2(n_m^T u_m)u_m = (I - 2u_m u_m^T)n_m = \left(\prod_{i=1}^{m} I - 2u_i u_i^T\right) n_1.$$

Second, we require $\mu_j > 0$ for every $j \in \{1, \ldots, m\}$, where

$$\mu_j = 2n_j^T u_j = 2u_j^T \left(\prod_{i=1}^{j-1} I - 2u_i u_i^T \right) n_1.$$

Here we can replace the strict inequality by $\geqslant$. The reason for this is the observation made earlier (see page 45), that the closed polygonal line γ is unique up to homothetic transformations. So, if a solution of the SOCP yields $\mu_i = 0$ for some i, then every solution yields $\mu_i = 0$ and no regular closed EBT $(q_1, \ldots, q_m)$ with $q_j \in F_j$ for $j \in \{1, \ldots, m\}$ exists.

For the objective of the SOCP we note that

$$\sum_{j=1}^{m} \mu_j = \sum_{j=1}^{m} 2n_j^T u_j = \sum_{j=1}^{m} 2u_j^T \left(\prod_{i=1}^{j-1} I - 2u_i u_i^T \right) n_1. \tag{4.6}$$

The vectors $n_1, \ldots, n_m$ are required to be vertices of $\mu\gamma + t$ for some $(\mu, t) \in S_\gamma$. Lemma 4.3.1 states that the only possible way for $n_1, \ldots, n_m$ to form a totally cyclic vector configuration is if μ is maximal. Hence, we would like to choose n_1 such that (4.6) is as large as possible. Thus, we obtain the following SOCP:

$$\begin{aligned}
\max \quad & \sum_{j=1}^{m} 2u_j^T \left(\prod_{i=1}^{j-1} I - 2u_i u_i^T \right) x \\
\text{s. t.} \quad & 2u_j^T \left(\prod_{i=1}^{j-1} I - 2u_i u_i^T \right) x \geqslant 0 \quad \forall\, j \in \{1, \ldots, m\} \\
& \left(\left(\prod_{i=1}^{m} I - 2u_i u_i^T \right) - I \right) x = 0 \\
& x \in \mathbb{R}^n,\ ||x|| \leqslant 1.
\end{aligned}$$

We note that this SOCP is not in primal standard form as introduced in Chapter 3.2.2 but we can bring it into primal standard form by introducing slack variables.

If there is an optimal solution x^*, we pick $n_1 = x^*$. Then it is easy to see that n_1 lies on S^{n-1}. First, we note that the optimal value is nonnegative because of the first constraint. In fact, if there is a regular closed EBT for the given choice of facets, then the optimal value is positive. Otherwise, an optimal value of 0 would imply $\mu_j = 0$ for every $j \in \{1, \ldots, m\}$. Thus, if we assume $||x^*|| = ||n_1|| < 1$ then $x^*/||x^*||$ would be a feasible solution as well but would have a greater objective value. This would contradict the optimality of x^*.

We now have a way to find n_1 and hence also $n_2, \ldots, n_m$. These vectors are unique by Lemma 4.3.2. This concludes our discussion on the third step of Algorithm 1 and we move on to the fourth step.

We proceed in a similar fashion as before to find the bouncing points $q_1, \ldots, q_m$ of a regular closed EBT. Equation (4.1) states that $q_{j+1} - q_j = \lambda_j n_j$ needs to hold for every $j \in \{1, \ldots, m\}$, where $\lambda_j > 0$ and $q_{m+1} := q_1$. Thus, just like $u_1, \ldots, u_m$, we have that

$n_1, \dots, n_m$ are linearly dependent:

$$0 = \sum_{j=1}^{m} (q_{j+1} - q_j) = \sum_{j=1}^{m} \lambda_j n_j.$$

We would like the $(n \times m)$-matrix $(n_1, \dots, n_m)$ to have rank $m - 1$. Then the closed polygonal line that we get by successively moving in the directions $n_1, \dots, n_m$ is unique up to homothetic transformations. Indeed it can be shown that this is the case if we make two assumptions. If for a given choice of facets $F_1, \dots, F_m$ one or both of these assumptions are violated, then we proceed with the next choice of facets.

First, we assume that the bouncing points $q_1, \dots, q_m$, which we are going to find in the remainder, belong to a regular closed EBT with minimal length. In particular, this includes the assumption that among all closed EBTs with minimal length there is one that is regular.

Second, we assume that the convex hull of $q_1, \dots, q_m$ has dimension $m - 1$. Note that if the first assumption is satisfied and if K is a strictly convex body with smooth boundary, then Theorem 4.2.2 implies that the second assumption is satisfied as well. However, in our case K is a polytope, i.e. neither strictly convex nor with smooth boundary.

Using these two assumptions we now show that the matrix that contains $n_1, \dots, n_m$ as columns has rank $m - 1$. We clearly have

$$\operatorname{rk}(n_1, \dots, n_m) \leqslant m - 1$$

because $n_1, \dots, n_m$ are linearly dependent. Due to the assumption that the convex hull of $q_1, \dots, q_m$ has dimension $m - 1$, we find that

$$q_1 - q_m,\ q_2 - q_m,\ \dots\ ,\ q_{m-1} - q_m$$

are linearly independent. Recall that $\operatorname{rk}(AB) = \operatorname{rk}(A)$ for $A \in \mathbb{R}^{n \times m}$ and $B \in \mathrm{Gl}_m(\mathbb{R})$. Thus, we get:

$$\begin{aligned}
m - 1 &= \operatorname{rk}\begin{pmatrix} q_1 - q_m & q_2 - q_m & \dots & q_{m-1} - q_m \end{pmatrix} \\
&= \operatorname{rk}\left(\begin{pmatrix} q_1 - q_m & q_2 - q_m & \dots & q_{m-1} - q_m \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & -1 \\ 0 & & & 1 \end{pmatrix} \right) \\
&= \operatorname{rk}\begin{pmatrix} q_1 - q_m & q_2 - q_1 & q_3 - q_2 & \dots & q_{m-1} - q_{m-2} \end{pmatrix} \\
&= \operatorname{rk}\begin{pmatrix} \lambda_m n_m & \lambda_1 n_1 & \lambda_2 n_2 & \dots & \lambda_{m-2} n_{m-2} \end{pmatrix} \\
&= \operatorname{rk}\begin{pmatrix} n_m & n_1 & n_2 & \dots & n_{m-2} \end{pmatrix} \\
&\leqslant \operatorname{rk}\begin{pmatrix} n_1 & n_2 & \dots & n_m \end{pmatrix}.
\end{aligned}$$

The second to last line of the equation holds since $\lambda_m, \lambda_1, \dots, \lambda_{m-2} > 0$. Thus, it follows that the closed polygonal line which we get by successively moving in the directions $n_1, \dots, n_m$ is unique up to homothetic transformations. In the following we let ξ be such a closed polygonal line, i.e. $\xi = (\xi_1, \dots, \xi_m)$ and

$$\xi_{j+1} - \xi_j = \lambda'_j n_j$$

with $\lambda'_j > 0$ for $j \in \{1, \dots, m\}$, where $\xi_{m+1} := \xi_1$. The task is now to find $\lambda > 0$ and $s \in \mathbb{R}^n$ such that $\lambda\xi_j + s \in F_j$ for every j. We can complete this task using an LP (see Chapter 3.2.1). For $j \in \{1, \dots, m\}$ let

$$H_j = \{x \in \mathbb{R}^n \colon u_j^T x = b_j\}$$

be the unique supporting hyperplane of K that contains F_j. Then we require

$$\begin{aligned}
\lambda\xi_j + s \in F_j &\iff \lambda\xi_j + s \in K \cap H_j \\
&\iff (\lambda\xi_j + s)^T u_j = b_j,\ (\lambda\xi_j + s)^T u_i \leqslant b_i \ \ \forall\, i \neq j \\
&\iff \begin{pmatrix} \xi_j^T u_j & u_j^T \end{pmatrix} \begin{pmatrix} \lambda \\ s \end{pmatrix} = b_j,\ \begin{pmatrix} \xi_j^T u_i & u_i^T \end{pmatrix} \begin{pmatrix} \lambda \\ s \end{pmatrix} \leqslant b_i \ \ \forall\, i \neq j.
\end{aligned}$$

Additionally, we would like to make sure that (if possible) the resulting closed EBT $(q_1, \dots, q_m)$ with $q_j = \lambda\xi_j + s$ is regular. Thus, we maximize the smallest slack of the form $b_i - (\lambda\xi_j^T u_i + u_i^T s)$ with $1 \leqslant i, j \leqslant m$ and $i \neq j$. This leads us to the following LP with $n + 2$ variables:

$$\begin{aligned}
\max \quad & \rho \\
\text{s. t.} \quad & \rho, \lambda \in \mathbb{R}_{\geqslant 0},\ s \in \mathbb{R}^n \\
& \begin{pmatrix} \xi_j^T u_j & u_j^T \end{pmatrix} \begin{pmatrix} \lambda \\ s \end{pmatrix} = b_j \ \ \forall\ j \in \{1, \dots, m\} \\
& \rho \leqslant b_i - \begin{pmatrix} \xi_j^T u_i & u_i^T \end{pmatrix} \begin{pmatrix} \lambda \\ s \end{pmatrix} \ \ \forall\, i, j \in \{1, \dots, m\},\ i \neq j.
\end{aligned}$$

Again we can bring this problem into dual standard form by introducing slack variables. Note that every solution of the LP satisfies $\lambda > 0$. If otherwise $\lambda = 0$, then the equality constraints state that $s \in H_j$ for every $j \in \{1, \dots, m\}$. This means all supporting hyperplanes that contain a facet of K intersect at the common point s, which is not possible if K is a polytope.

If the LP has an optimal solution (ρ^*, λ^*, s^*), then we let $q_j = \lambda^*\xi_j + s^*$ for $j \in \{1, \dots, m\}$. Moreover, if $\rho^* > 0$, then $(q_1, \dots, q_m)$ is a regular closed EBT that potentially has minimal length, meaning it fulfils the conditions in Theorem 4.2.1 and Theorem 4.2.2, i.e.:

$$\begin{aligned}
\dim(\operatorname{conv}\{q_1, \dots, q_m\}) &= m - 1, \\
\dim(\operatorname{cone}\{u_1, \dots, u_m\}) &= m - 1.
\end{aligned}$$

To ensure that we do not miss regular closed EBTs that potentially have minimal length, we prove that it suffices to find one such closed EBT per choice of $F_1, \dots, F_m$.

Theorem 4.3.3. *Let $F_1, \dots, F_m$ be facets of some full-dimensional polytope $K \subseteq \mathbb{R}^n$ and let $q_j, q'_j \in F_j$ for every $j \in \{1, \dots, m\}$. Assume $q = (q_1, \dots, q_m)$ and $q' = (q'_1, \dots, q'_m)$ are regular closed EBTs. Furthermore, assume that the conic hull of the one-dimensional normal cones $N_K(q_1), \dots, N_K(q_m)$ is a linear space with dimension $m - 1$. Then:*

$$\ell_{B_1(0)}(q) = \ell_{B_1(0)}(q').$$

Proof. According to Theorem 4.1.2 there are $n_1, \dots, n_m, n'_1, \dots, n'_m \in S^{n-1}$ such that for every $j \in \{1, \dots, m\}$:

$$\begin{cases} q_{j+1} - q_j = \lambda_j n_j,\ \lambda_j \geqslant 0, \\ n_{j+1} - n_j \in -N_K(q_{j+1}) \end{cases} \quad \text{and} \quad \begin{cases} q'_{j+1} - q'_j = \lambda'_j n'_j,\ \lambda'_j \geqslant 0, \\ n'_{j+1} - n'_j \in -N_K(q'_{j+1}) \end{cases}$$

where $q_{m+1} := q_1$, $n_{m+1} := n_1$ and $n'_{m+1} := n'_1$. As we discuss earlier in this chapter, $n_1, \dots, n_m$ are unique and only depend on $N_K(q_1), \dots, N_K(q_m)$, due to the assumption that the conic hull of $N_K(q_1), \dots, N_K(q_m)$ is a linear space with dimension $m - 1$. Therefore, we have $n'_j = n_j$ for $j \in \{1, \dots, m\}$. Identity (4.2) now states that

$$\ell_{B_1(0)}(q) = \sum_{j=1}^{m} (q_{j+1} - q_j)^T n_j,$$
$$\ell_{B_1(0)}(q') = \sum_{j=1}^{m} (q'_{j+1} - q'_j)^T n_j.$$

Thus, we get:

$$\begin{aligned} \ell_{B_1(0)}(q) - \ell_{B_1(0)}(q') &= \sum_{j=1}^{m} \left[(q_{j+1} - q_j)^T n_j - (q'_{j+1} - q'_j)^T n_j \right] \\ &= \sum_{j=1}^{m} \left[(q_{j+1} - q'_{j+1})^T n_j + (q'_j - q_j)^T n_j \right] \\ &= \sum_{j=1}^{m} (q_{j+1} - q'_{j+1})^T n_j + \sum_{j=1}^{m} (q'_{j+1} - q_{j+1})^T n_{j+1} \\ &= \sum_{j=1}^{m} (q'_{j+1} - q_{j+1})^T (n_{j+1} - n_j). \end{aligned}$$

For every $j \in \{1, \dots, m\}$ we let u_j be the unique unit vector in $N_K(q_{j+1})$ and we let

$$H_{j+1} = \{y \in \mathbb{R}^n : u_j^T x = b_{j+1}\}$$

be the supporting hyperplane of K that contains F_{j+1}, where $F_{m+1} := F_1$. On the one hand, we have

$$u_j^T (q'_{j+1} - q_{j+1}) = u_j^T q'_{j+1} - u_j^T q_{j+1} = b_{j+1} - b_{j+1} = 0.$$

On the other hand, $n_{j+1} - n_j \in -N_K(q_{j+1})$. Thus, $n_{j+1} - n_j$ is a multiple of u_j. This yields $\ell_{B_1(0)}(q) - \ell_{B_1(0)}(q') = 0$. □

We conclude that if there is a closed EBT $(q_1, \dots, q_m)$ with minimal length that is regular and such that

$$\dim(\operatorname{conv}\{q_1, \dots, q_m\}) = m - 1,$$
$$\dim(\operatorname{cone}\{u_1, \dots, u_m\}) = m - 1,$$

then Algorithm 1 returns its length. If there is no closed EBT with these properties, the algorithm returns an upper bound on the minimal length. Together with Theorem 4.1.3 we have an algorithm that computes an upper bound on $c_{\mathrm{EHZ}}(K \times B_1(0))$.

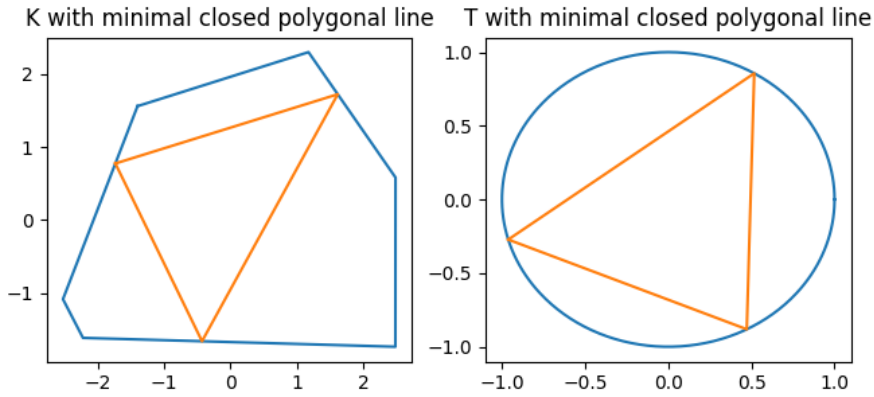


Figure 4.6: An exemplary output of Algorithm 1. The input is a randomly generated polygon K. K and the two-dimensional Euclidean unit ball are depicted in blue. The calculated closed polygonal lines ((q_1, q_2, q_3) on the left and (n_1, n_2, n_3) on the right) are depicted in orange.

For each choice of $F_1, \ldots, F_m$ Algorithm 1 needs to solve the following tasks: Calculate the rank of an $(n \times m)$-matrix, solve an SOCP with $n+m+1$ constraints and $n+1$ variables, solve an $(n+1) \times m$ system of linear equations and solve an LP with mf constraints and $n+2$ variables, where f is the number of facets of K. All these tasks are solvable in polynomial time with respect to the dimension n and the number of facets of K. However, there are $\sum_{j=2}^{n+1} \binom{f}{j} j!$ possibilities to choose at least 2 but at most $n+1$ facets, respecting their order. We can slightly reduce this number since a circular shift of the chosen facets $F_1, \ldots, F_m$ will yield a similar but shifted result. This leaves us with $\sum_{j=2}^{n+1} \binom{f}{j} (j-1)!$ possibilities. The calculations for each of these possibilities are independent of each other. Therefore, we utilize parallel computing to accelerate the execution of the algorithm.

Table 6.1 in the appendix collects the running time of Algorithm 1 for some instances where the billiard table is a polytope K of dimension 2, 3 or 4 which we generate in the following way. First, we choose some normally distributed random vectors. We scale each of these vectors by some scalar between 1 and 3 and we decrease the length of this range if the amount of random vectors becomes too large. Afterwards, we receive K as the convex hull of these vectors. Instead of a total running time the table shows the time needed to compute a suitable regular closed EBT with 2, 3, 4 and 5 bouncing points respectively. The table suggests that the calculations for m bouncing points with $m < n+1$ terminate quickly. The reason for this is that many iterations are rejected early when the rank of $(u_1, \ldots, u_m)$ is checked. All calculations have been done on a Dell Latitude E6530 laptop with Intel Core i7-3520M processor, 2.9 GHz (capable of running four threads). We use the programming language Python [96] and mainly utilize the NumPy library [48]. Furthermore, we solve the LPs and SOCPs using the respective default solvers of the software Mosek [8], which

employ interior-point methods. Throughout this thesis we use the same hard- and software unless specified otherwise.

4.3.2 Governing the bouncing rule with polytopes

In the remainder of Chapter 4 we pay attention to the Minkowski setting. Thus, we let T be a convex set that is not necessarily the unit ball. Most of the results in Chapter 4.1 and 4.2 require T to be strictly convex and some of them also require that the boundary of T is smooth. Our next goal is to recover the strategy from the end of Chapter 4.1 for the case where both $K, T \subseteq \mathbb{R}^n$ are full-dimensional polytopes. The reason for this is that polytopes allow for an easy implementation via their vertices or their facets. Besides, we can approximate every convex body with polytopes.

Throughout Chapter 4.3.2 we are concerned with sequences in $\mathbb{R}^n$ as well as with sequences of subsets of $\mathbb{R}^n$. If we say that a sequence in $\mathbb{R}^n$ converges, then we mean convergence with respect to the standard Euclidean metric. On the other hand, if we say that a sequence of subsets of $\mathbb{R}^n$ converges, then we mean convergence with respect to the Hausdorff metric d_H unless specified otherwise. We recall the definition of the Hausdorff metric as given in Chapter 2.3:

$$d_H(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B}||a-b||,\ \sup_{b\in B}\inf_{a\in A}||a-b||\right\},\ \ A,B\subseteq\mathbb{R}^n.$$

Theorem 4.2.4 states that if we take a sequence $(T_i)_{i\in\mathbb{N}}$ of strictly convex bodies in $\mathbb{R}^n$ that converges to T, then we can find a subsequence $(T_{i_j})_{j\in\mathbb{N}}$ such that the sequence of closed (K,T_{i_j})-MBTs with minimal $\ell_{T_{i_j}}$-length converges to a closed polygonal line q^* such that the set of vertices of q^* lies in

$$F(K) = \{M\subseteq\mathbb{R}^n\colon \nexists\, t\in\mathbb{R}^n,\ M+t\subseteq \operatorname{int} K\}$$

and such that the ℓ_T-length of q^* is minimal among all closed polygonal lines with this property. By Theorem 4.2.3 the set of vertices of every closed (K,T_{i_j})-MBT is contained in

$$F_{n+1}(K) = \{M\subseteq\mathbb{R}^n\colon |M|\leqslant n+1,\ M\in F(K)\}.$$

Moreover, since the vertices of a closed MBT lie on the boundary of K, they are contained in some ball with sufficiently large radius. Using a standard compactness argument we conclude that the set of vertices of q^* lies in $F_{n+1}(K)$ as well. Together with Theorem 4.1.3 and Lemma 2.3.3, we have

$$c_{\mathrm{EHZ}}(K\times T) = \min\{\ell_T(q) : q=(q_1,\dots,q_k) \text{ with } \{q_1,\dots,q_k\}\in F_{n+1}(K),\ k\geqslant 2\}.$$

The algorithm that we discuss in Chapter 4.3.3 expects two full-dimensional polygons $K,T\subseteq\mathbb{R}^2$ as input. It finds closed polygonal lines $(q_1,\dots,q_m)$ and $(p_1,\dots,p_m)$ with $q_j\in\partial K$ and $p_j\in\partial T$ for $j\in\{1,\dots,m\}$ and $m\in\{2,3\}$ such that (4.1) holds and q has minimal ℓ_T-length among all these lines. The following theorem proves that this suffices to compute the EHZ-capacity of $K\times T$.

Theorem 4.3.4. *Let $K, T \subseteq \mathbb{R}^n$ be convex bodies. Furthermore, let q be a closed polygonal line such that the set of vertices of q lies in $F_{n+1}(K)$ and such that q has minimal ℓ_T-length among all closed polygonal lines with this property. Then there exists a closed polygonal line $q^* = (q_1, \dots, q_m)$ with*

$$q^* \in \operatorname{argmin}\{\ell_T(\widetilde{q}) : \widetilde{q} \text{ is a closed polygonal line and there is a closed polygonal line } p \text{ so that the pair } (\widetilde{q}, p) \text{ fulfils (4.1)}\}$$

such that $m \in \{2, \dots, n+1\}$ and $\ell_T(q^*) = \ell_T(q)$.

Proof. We start by letting $(T_i)_{i\in\mathbb{N}}$ be a sequence of strictly convex bodies in $\mathbb{R}^n$ that converges to T. Furthermore, we let this sequence be an inner approximation, i.e.:

$$T_i \subseteq T \text{ for all } i \in \mathbb{N}.$$

Applying the same reasoning as before (i.e. apply Theorem 4.2.4 and Theorem 4.2.3, take a subsequence and make a standard compactness argument) we find a converging sequence $(q^i)_{i\in\mathbb{N}}$ of closed (K, T_i)-MBTs with minimal ℓ_{T_i}-length. Furthermore, this sequence converges to some closed polygonal line $q^* \in F_{n+1}(K)$ with minimal ℓ_T-length. In particular, we have $\ell_T(q^*) = \ell_T(q)$. It remains to show that there is a closed polygonal line p^* such that (q^*, p^*) fulfils (4.1).

Since q^i is a closed (K, T_i)-MBT, by Theorem 4.1.2 we can find a dual closed MBT p^i in T_i such that

$$\begin{cases} q^i_{j+1} - q^i_j \in N_{T_i}(p^i_j), \\ p^i_{j+1} - p^i_j \in -N_K(q^i_{j+1}) \end{cases}$$

holds for all $j \in \{1, \dots, m\}$ and for all $i \in \mathbb{N}$. Here, $q^i_{m+1} := q^i_1$ and $p^i_{m+1} := p^i_1$. We now consider the sequence $(p^i_j)_{i\in\mathbb{N}}$ for some $j \in \{1, \dots, m\}$. Every element of this sequence lies in the convex body T. We recall that a convex body is a compact set. Thus, the Bolzano-Weierstraß theorem [15] implies that there is a converging subsequence of $(p^i_j)_{i\in\mathbb{N}}$. By successively repeating this process for every $j \in \{1, \dots, m\}$ we find that the sequence $((p^i_1, \dots, p^i_m))_{i\in\mathbb{N}}$ has a converging subsequence. We now switch to this subsequence while keeping the notation $((p^i_1, \dots, p^i_m))_{i\in\mathbb{N}}$ for the sake of simplicity. We write

$$p^*_j := \lim_{i\to\infty} p^i_j,$$

for every $j \in \{1, \dots, m\}$. We have $p^*_j \in \partial T$ because on the one hand p^i_j is contained in ∂T_i and on the other hand $(T_i)_{i\in\mathbb{R}^n}$ converges to T. Next, we show that (by switching to a subsequence if necessary)

$$\lim_{i\to\infty} N_{T_i}(p^i_j) \subseteq N_T(p^*_j),$$

for every $j \in \{1, \dots, m\}$. We need to specify what these limits mean since the Hausdorff metric is inconvenient if we talk about cones. More precisely, if we have two different cones, then the Hausdorff distance between them is always infinite. Instead we say that

a sequence of cones $(C_i)_{i\in\mathbb{N}}$ converges to some cone C, if the sequence $(C_i \cap B_1(0))_{i\in\mathbb{N}}$ converges to $C \cap B_1(0)$ with respect to the Hausdorff metric.

We fix $j \in \{1,\dots,m\}$ and let

$$u_j \in \lim_{i\to\infty} N_{T_i}(p_j^i).$$

Then there is a sequence $(u_j^i)_{i\in\mathbb{N}}$ with $u_j^i \in N_{T_i}(p_j^i)$ that converges to u_j. By the definition of the normal cone we have

$$(u_j^i)^T(z - p_j^i) \leqslant 0,$$

for all $z \in T_i$ and every $i \in \mathbb{N}$. In particular, for every $z^* \in T$ we can take a sequence $(z_i)_{i\in\mathbb{N}}$ with $z_i \in T_i$ that converges to z^* and get

$$(u_j^i)^T(z_i - p_j^i) \leqslant 0.$$

Therefore, by continuity we have

$$(u_j)^T(z^* - p_j^*) = \lim_{i\to\infty} (u_j^i)^T(z_i - p_j^i) \leqslant 0.$$

It follows that $u_j \in N_T(p_j^*)$ and thus

$$\lim_{i\to\infty} N_{T_i}(p_j^i) \subseteq N_T(p_j^*).$$

A similar argument proves that $\lim_{i\to\infty} N_K(q_{j+1}^i) \subseteq N_K(q_{j+1}^*)$. In total, we get:

$$\begin{aligned} q_{j+1}^* - q_j^* &= \lim_{i\to\infty} \left(q_{j+1}^i - q_j^i\right) \in \lim_{i\to\infty} N_T(p_j^i) \subseteq N_T(p_j^*), \\ p_{j+1}^* - p_j^* &= \lim_{i\to\infty} \left(p_{j+1}^i - p_j^i\right) \in \lim_{i\to\infty} -N_K(q_{j+1}^i) \subseteq -N_K(q_{j+1}^*) \end{aligned}$$

for all $j \in \{1,\dots,m\}$, where $q_{m+1}^* := q_1^*$ and $p_{m+1}^* := p_1^*$. So, the pair of closed polygonal lines (q^*, p^*) with $p^* = (p_1^*,\dots,p_m^*)$ fulfils (4.1). □

Now, we know that to compute $c_{\mathrm{EHZ}}(K \times T)$, where $K, T \subseteq \mathbb{R}^n$ are convex bodies, it is sufficient to find closed polygonal lines q, p with m vertices, $m \in \{2,\dots,n+1\}$, such that the pair (q,p) satisfies (4.1) and such that q has minimal ℓ_T-length. If T is strictly convex with smooth boundary and if we consider $m = n+1$, then by Theorem 4.2.1 we can demand that the vertices of q are smooth boundary points of K. We can extend the proof of Theorem 4.3.4 to obtain a similar result if T is an arbitrary convex body and if K is a polytope.

Theorem 4.3.5. *Let $K \subseteq \mathbb{R}^n$ be a full-dimensional polytope and let $T \subseteq \mathbb{R}^n$ be a convex body. Furthermore, let*

$$\begin{aligned} Q = \{\widetilde{q}\colon\ &\widetilde{q} \text{ is a closed polygonal line and there is a closed} \\ &\text{polygonal line } p \text{ so that the pair } (\widetilde{q}, p) \text{ fulfils } (4.1)\}. \end{aligned}$$

Assume that every closed polygonal line $q \in Q$ with minimal ℓ_T-length has exactly $n+1$ vertices. Then there is a closed polygonal line $q^* := (q_1^*, \dots, q_{n+1}^*) \in Q$ with minimal ℓ_T-length and a closed polygonal line $p^* := (p_1^*, \dots, p_{n+1}^*)$ such that (q^*, p^*) fulfils (4.1) *and for every $j \in \{1, \dots, n+1\}$ we have*

$$p_j^* - p_{j-1}^* \in -N_K(z_j),$$

where $p_0^* := p_{n+1}^*$ and z_j is a smooth boundary point of K so that z_j and q_j^* lie in a common facet of K.

Proof. We let $(T_i)_{i\in\mathbb{N}}$ be a sequence of strictly convex bodies in $\mathbb{R}^n$ with smooth boundaries such that this sequence converges to T. Like in the proof of Theorem 4.3.4 we find a sequence of closed (K, T_i)-MBTs $((q_1^i, \dots, q_{n+1}^i))_{i\in\mathbb{N}}$ and a sequence of dual closed MBTs $((p_1^i, \dots, p_{n+1}^i))_{i\in\mathbb{N}}$ such that the pair of closed polygonal lines (q^*, p^*) fulfils (4.1), where

$$q_j^* = \lim_{i\to\infty} q_j^i,$$
$$p_j^* = \lim_{i\to\infty} p_j^i$$

for every $j \in \{1, \dots, n+1\}$. Moreover, q^* has minimal ℓ_T-length among all closed polygonal lines in Q. Theorem 4.2.1 implies that the bouncing points of $q^i = (q_1^i, \dots, q_{n+1}^i)$ are smooth boundary points of K for every $i \in \mathbb{N}$. In other words, $N_K(q_j^i)$ is one-dimensional for every $i \in \mathbb{N}$ and every $j \in \{1, \dots, n+1\}$. If the points $q_1^*, \dots, q_{n+1}^*$ are smooth boundary points of K as well, then the statement follows immediately since

$$p_j^* - p_{j-1}^* \in -N_K(q_j^*)$$

holds for every $j \in \{1, \dots, n+1\}$. Therefore, we assume that q_j^* is not a smooth boundary point of K for some j. In particular, $N_K(q_j^*)$ is not one-dimensional. We note that $(q_j^i)_{i\in\mathbb{N}}$ and $(p_j^i)_{i\in\mathbb{N}}$ are Cauchy sequences. Hence for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that for every $i, k \geqslant N$:

$$\begin{aligned} |(-\mu_{i,j}u_j^i) - (-\mu_{k,j}u_j^k)| &= |(p_j^i - p_{j-1}^i) - (p_j^k - p_{j-1}^k)| \\ &\leqslant |p_j^i - p_j^k| + |p_{j-1}^i - p_{j-1}^k| \\ &\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

where u_j^i denotes the unique unit vector in $N_K(q_j^i)$ and $\mu_{i,j}, \mu_{k,j} \geqslant 0$ are some scalars as in (4.1). Here, $p_0^i := p_{n+1}^i$ and $p_0^k := p_{n+1}^k$. Furthermore, we have

$$\lim_{i\to\infty} -\mu_{i,j}u_j^i = \lim_{i\to\infty} \left(p_j^i - p_{j-1}^i\right) \neq 0.$$

Otherwise, $p_j^* = p_{j-1}^*$ which would imply that

$$\widehat{q} = (q_1^*, \dots, q_{j-1}^*, q_{j+1}^*, \dots, q_{n+1}^*),$$
$$\widehat{p} = (p_1^*, \dots, p_{j-1}^*, p_{j+1}^*, \dots, p_{n+1}^*)$$

satisfy (4.1) and that $\ell_T(\widehat{q}) \leqslant \ell_T(q^*)$ holds. This would contradict the assumption that every closed polygonal line in Q with minimal ℓ_T-length has exactly $n+1$ vertices.

Since K is a polytope, it has only finitely many facets. Therefore, there are finitely many different outer normal cones of the form $N_K(z)$ with $z \in \partial K$. Taking into account that q_j^i is a smooth boundary point of K for every $i \in \mathbb{N}$, this means that the set

$$\left\{u_j^i : i \in \mathbb{N}\right\}$$

is finite for every $j \in \{1,\dots,n+1\}$. More precisely, its cardinality is at most the number of facets of K. So, if multiples of u_j^i and u_j^k become arbitrary close and do not approach 0 as i and k increase, we conclude that there is $N' \in \mathbb{N}$ such that $u_j^i = u_j^k$ holds for $i,k \geqslant N'$. Therefore, we can denote $u_j := u_j^i$ for $i \geqslant N'$. From this follows:

$$p_j^* - p_{j-1}^* = \lim_{i\to\infty}\left(p_j^i - p_{j-1}^i\right) = \lim_{i\to\infty} -\mu_{i,j}u_j.$$

Thus, the claim follows if we repeat this argument for every $j \in \{1,\dots,n+1\}$. More precisely, we pick $z_j = q_j^*$ if q_j^* is a smooth boundary point of K and otherwise we pick $z_j = u_j$. □

Theorem 4.3.4 and Theorem 4.3.5 motivate the following strategy to compute the EHZ capacity. Given two full-dimensional polytopes K, T we search for two closed polygonal lines $q = (q_1,\dots,q_m)$ and $p = (p_1,\dots,p_m)$ such that $q_j \in \partial K$, $p_j \in \partial T$ for every $j \in \{1,\dots,m\}$ and such that (4.1) is fulfilled. We do so for every $m \in \{2,\dots,n+1\}$. If $m = n+1$ we demand that q and p satisfy the following system which is more restrictive than (4.1):

$$\begin{cases} q_{j+1} - q_j \in N_T(p_j), \\ p_{j+1} - p_j \in -N_K(z_{j+1}), \end{cases}$$

for some smooth boundary points $z_1,\dots,z_m \in \partial K$ such that z_j and q_j are contained in a common facet of K for every $j \in \{1,\dots,m\}$. Here, $q_{m+1} := q_1$, $p_{m+1} := p_1$ and $z_{m+1} := z_1$. Together, the observations in Chapter 4.3.2 ensure that the minimal ℓ_T-length among all found closed polygonal lines q equals $c_{\mathrm{EHZ}}(K \times T)$.

4.3.3 The Minkowski setting in dimension 4

Next, we elaborate on the implementation of our strategy to compute the EHZ capacity in the Minkowski setting. Before we can describe the algorithm in more detail, we need a statement similar to Theorem 4.3.3. More precisely, we show that if we are given two pairs of closed polygonal lines (q,p) and (q',p'), where the vertices q_j, q_j' lie in the relative interior of a common face of K and the vertices p_j, p_j' lie in the relative interior of a common face of T for every $j \in \{1,\dots,m\}$ and such that both pairs fulfil (4.1), then q and q' have the same ℓ_T-length.

Theorem 4.3.6. *Let $K, T \subseteq \mathbb{R}^n$ be nonempty, convex sets, let $F_1, \ldots, F_m$ be faces of K and let $G_1, \ldots, G_m$ be faces of T for some $m \geqslant 2$. Furthermore, let $q = (q_1, \ldots, q_m)$, $q' = (q'_1, \ldots, q'_m)$ be closed polygonal lines with vertices on ∂K. Assume there are closed polygonal lines $p = (p_1, \ldots, p_m)$, $p' = (p'_1, \ldots, p'_m)$ with vertices on ∂T such that both (q, p) and (q', p') fulfil* (4.1)*. Moreover, assume for each $j \in \{1, \ldots, m\}$ that $q_j, q'_j \in \operatorname{relint} F_j$ and $p_j, p'_j \in \operatorname{relint} G_j$ (where we apply the convention that the relative interior of a single point is the point itself). Then $\ell_T(q) = \ell_T(q')$.*

Proof. We start the proof by stating a simple fact. If F is a face of a polytope P with $\dim(F) \geqslant 1$ and $y_1, y_2 \in F$, then

$$(y_1 - y_2)^T v = 0 \quad \forall v \in N_P(z) \tag{4.7}$$

holds for every $z \in \operatorname{relint} F$. To see this, consider the affine hull of F and shift it such that it is a linear space, i.e. such that it contains the origin. Then the vector $y_1 - y_2$ is an element of this space and $N_P(z)$ with $z \in \operatorname{relint} F$ is contained in the corresponding orthogonal space. Note that (4.7) also holds if F is a vertex because then $y_1, y_2 \in F$ implies $y_1 = y_2$. Now recall (4.1):

$$\begin{aligned} q_{j+1} - q_j &\in N_T(p_j), & q'_{j+1} - q'_j &\in N_T(p'_j), \\ p_{j+1} - p_j &\in -N_K(q_{j+1}), & p'_{j+1} - p'_j &\in -N_K(q'_{j+1}), \end{aligned}$$

where $q_{m+1} := q_1$, $p_{m+1} := p_1$, $q'_{m+1} := q'_1$ and $p'_{m+1} := p'_1$. The following calculation completes the proof:

$$\begin{aligned} \ell_T(q') &= \sum_{j=1}^{m} (q'_{j+1} - q'_j)^T p'_j \\ &= \sum_{j=1}^{m} \left[(q'_{j+1} - q'_j)^T p'_j + (q_{j+1} - q_j)^T p'_j - (q_{j+1} - q_j)^T p'_j \right] \\ &= \sum_{j=1}^{m} (q_{j+1} - q_j)^T p'_j + \sum_{j=1}^{m} (q_j - q'_j)^T p'_j - \sum_{j=1}^{m} (q_{j+1} - q'_{j+1})^T p'_j \\ &= \sum_{j=1}^{m} (q_{j+1} - q_j)^T p'_j + \sum_{j=1}^{m} (q_j - q'_j)^T p'_j - \sum_{j=1}^{m} (q_j - q'_j)^T p'_{j-1} \\ &= \sum_{j=1}^{m} (q_{j+1} - q_j)^T p'_j + \sum_{j=1}^{m} (q_j - q'_j)^T (p'_j - p'_{j-1}) = \sum_{j=1}^{m} (q_{j+1} - q_j)^T p'_j \\ &= \sum_{j=1}^{m} (q_{j+1} - q_j)^T p_j + \sum_{j=1}^{m} (q_{j+1} - q_j)^T (p'_j - p_j) = \sum_{j=1}^{m} (q_{j+1} - q_j)^T p_j = \ell_T(q), \end{aligned}$$

where $p'_0 := p'_m$. In the last and second to last lines of this equation we use the following identity, which follows from (4.7):

$$(q_j - q'_j)^T (p'_{j-1} - p'_j) = 0 = (q_{j+1} - q_j)^T (p'_j - p_j). \qquad \square$$

Theorem 4.3.6 and the observations from Chapter 4.3.2 suggest the following two algorithms to find closed MBTs with two or three bouncing points that have minimal ℓ_T-length. Together, these algorithms provide a way to compute $c_{\mathrm{EHZ}}(K \times T)$, where $K, T \subseteq \mathbb{R}^2$ are polygons. The input for both algorithms are the faces, i.e. vertices and edges, of K and T. In the following we let

$$\mathcal{F}(P) := \{F \subseteq P \colon F \text{ is a vertex or an edge of } P\}$$

for every polygon P.

Algorithm 2 Shortest closed MBT with two bouncing points

1: **for** every choice of $F_1, F_2 \in \mathcal{F}(K)$ with $F_1 \neq F_2$ and $G_1, G_2 \in \mathcal{F}(T)$ with $G_1 \neq G_2$ **do**

2: Let $N_{K,j} = N_K(y)$ for some $y \in \operatorname{relint} F_j$ and let $N_{T,j} = N_T(z)$ for some $z \in \operatorname{relint} G_j$ and for $j \in \{1, 2\}$. Find points $q_j \in F_j$ and $p_j \in G_j$ such that

$$\begin{cases} q_{j+1} - q_j \in N_{T,j}, \\ p_{j+1} - p_j \in -N_{K,j+1}, \end{cases} \tag{4.8}$$

where $q_3 := q_1$ and $p_3 := p_1$.

3: Calculate $\ell_T((q_1, q_2))$ and store this value if it is the smallest value found so far.

4: **end for**

Algorithm 3 Shortest closed MBT with three bouncing points

1: **for** every choice of three different edges F_1, F_2, F_3 of K **do**

2: Let u_j be the unique unit vector in $N_K(y)$ for some $y \in \operatorname{relint} F_j$. Construct a closed polygonal line γ by successively moving in the directions $-u_1, -u_2, -u_3$. Translate and scale γ with a positive factor such that its vertices lie on the boundary of T. Let p_1, p_2, p_3 be these vertices, i.e.:

$$p_{j+1} - p_j = -\mu_{j+1} u_{j+1}$$

with some suitable $\mu_j \in \mathbb{R}_{\geqslant 0}$ for every $j \in \{1, 2, 3\}$, $p_4 := p_1$, $u_4 := u_1$ and $\mu_4 := \mu_1$.

3: **for** every choice of unit vectors $w_j \in N_T(p_j)$ for every $j \in \{1, 2, 3\}$ **do**

4: Construct a closed polygonal line $\xi = (\xi_1, \ldots, \xi_m)$ by successively moving in the directions w_1, w_2, w_3. Translate and scale ξ with a positive factor to get a closed polygonal line $q = (q_1, q_2, q_3)$ with $q_j \in F_j$ for every $j \in \{1, 2, 3\}$. In particular:

$$q_{j+1} - q_j = \lambda_j w_j$$

with some suitable $\lambda_j \in \mathbb{R}_{\geqslant 0}$ for every $j \in \{1, 2, 3\}$ and $q_4 := q_1$.

5: Calculate $\ell_T(q)$ and store this value if it is the smallest value found so far.

6: **end for**

7: **end for**

We proceed to discuss Algorithm 2. For each choice of $F_1, F_2 \in \mathcal{F}(K)$ and for each choice of $G_1, G_2 \in \mathcal{F}(T)$ we compute closed polygonal lines $q = (q_1, q_2)$ and $p = (p_1, p_2)$ that fulfil (4.1) such that $q_j \in F_j$ and $p_j \in G_j$ for $j \in \{1, 2\}$. Note that Theorem 4.3.6 implies that Algorithm 2 considers for each

$$r \in \{\ell_T(\bar{q})\colon \bar{q} \text{ is a closed MBT with two bouncing points}\}$$

at least one closed MBT with ℓ_T-length r if we ask for $q_j \in \operatorname{relint} F_j$ and $p_j \in \operatorname{relint} G_j$ in each iteration. We allow $q_j \in \partial F_j$ and $p_j \in \partial G_j$ as well because we can extend the proof of Theorem 4.3.6 to the case where $q_j \in F_j$ and $p_j \in G_j$ such that (4.8) holds.

To reduce the running time we check whether $N_K(q_1) \cap -N_K(q_2)$ and $N_T(p_1) \cap -N_T(p_2)$ are nonempty before we start the calculation. The reason for this is that the existence of a pair (q, p) of closed polygonal lines that satisfy (4.1) implies:

$$\begin{aligned} -N_K(q_2) \ni p_2 - p_1 = -(p_1 - p_2) \in N_K(q_1),\\ N_T(p_1) \ni q_2 - q_1 = -(q_1 - q_2) \in -N_T(p_2). \end{aligned}$$

Note that the normal cones $N_K(q_j), N_T(p_j)$ only depend on F_j, G_j. So, in the following we can assume that these intersections are indeed nonempty. The goal is now to calculate a pair of suitable closed polygonal lines (q, p) if possible. To this end, it is helpful to distinguish F_j, G_j into edges and vertices. We consider the following cases:

1) F_1, F_2, G_1, G_2 are vertices.

2) F_1, F_2 are vertices and among G_1, G_2 there is at least one edge.

3) Among F_1, F_2 and among G_1, G_2 there is at least one edge.

All remaining cases can be covered by switching the roles of K and T.

The first case is easy. If F_1, F_2, G_1, G_2 are vertices, the resulting closed polygonal lines are unique and we can check (4.1) directly.

We start the investigation of the second case by assuming that both G_1 and G_2 are edges. Then $N_T(p_1)$ and $N_T(p_2)$ are one-dimensional cones. Let w_j be the unique unit vector in $N_T(p_j)$ for every $j \in \{1, 2\}$. We can verify whether $q_2 - q_1 \in N_T(p_1)$ holds by checking whether w_1 is a positive multiple of $q_2 - q_1$. If this is the case, then $q_1 - q_2 \in N_T(p_2)$ follows directly since we assume that $N_T(p_1) \cap -N_T(p_2)$ is nonempty. Alternatively, we can check whether w_2 is a positive multiple of $q_1 - q_2$ and get $q_2 - q_1 \in N_T(p_1)$ for free. It remains to solve the following problem:

$$\begin{aligned} &\text{Find } p_1, p_2 \text{ such that:}\\ &\qquad p_1 \in G_1\,,\ p_2 \in G_2,\\ &\qquad p_2 - p_1 \in -N_K(q_2),\\ &\qquad p_1 - p_2 \in -N_K(q_1). \end{aligned}$$

We can express the constraints by linear equations and inequalities. For this recall the definition of the outer normal cone of a convex set C at $z \in \partial C$:

$$N_C(z) = \{v\colon v^T(y - z) \leqslant 0 \ \ \forall\, y \in C\}.$$

If C is a polytope, it is sufficient to demand that $v^T(y-z) \leqslant 0$ for every vertex y of C. Therefore, we can model the membership of $N_C(z)$ with finitely many linear inequalities. Altogether, we can find suitable points p_1, p_2 using LP techniques. We use the same approach if either G_1 or G_2 is a vertex. In this case the LP remains unchanged except for the fact that we replace one of the two variable vectors by a constant vector.

In the third case we start with the assumption that F_1, F_2, G_1, G_2 are edges. Then all relevant normal cones are one-dimensional and we let

$$u_j \in N_K(q_j)\,,\ w_j \in N_T(p_j)$$

be unit vectors for every $j \in \{1,2\}$. Now we solve the following problem:

$$\begin{aligned}
&\text{Find } q_1, q_2, p_1, p_2, \alpha_1, \alpha_2 \text{ such that:}\\
&\quad q_1 \in F_1\,,\ q_2 \in F_2\,,\ p_1 \in G_1\,,\ p_2 \in G_2,\\
&\quad \alpha_1, \alpha_2 \geqslant 0,\\
&\quad q_2 - q_1 = \alpha_1 w_1,\\
&\quad p_2 - p_1 = -\alpha_2 u_2.
\end{aligned}$$

Similar to the previous case, this problem is an LP. Note that the last two constraints suffice to imply (4.1) since $N_T(p_1) \cap -N_T(p_2)$ and $N_K(q_1) \cap -N_K(q_2)$ are nonempty. If there are vertices among F_1, F_2, G_1, G_2, for instance if G_1 is a vertex, then the LP has to be changed in two places. First, like in the second case, we replace the corresponding variable vector, here p_1, with a constant vector. Second, if G_1 is a vertex, then the normal cone $N_T(p_1)$ is no longer one-dimensional and the definition of w_1 does not make sense any more. However, in this case G_2 is an edge and we replace the constraint $q_2 - q_1 = \alpha_1 w_1$ with $q_1 - q_2 = \alpha_1 w_2$. We apply the same reasoning if F_1 or F_2 is a vertex.

There may be multiple ways to choose (p,q) for a given choice of F_1, F_2, G_1, G_2. If this is the case, Algorithm 2 chooses (if possible) q such that $N_K(q_j)$ is one-dimensional for $j \in \{1,2\}$ (or equivalently such that q_1, q_2 are not vertices of K). We achieve this in the following manner. If F_1 or F_2 is a vertex, then the resulting closed polygonal line q always contains a vertex of K. So, we assume both F_1 and F_2 are edges. If q_1 and q_2 are smooth boundary points, i.e. lie in the relative interior of F_1 or F_2, then there is nothing to do. Otherwise, we denote

$$\begin{aligned}
N_K(q_1) &= \operatorname{cone}\{u\} = \{\beta u \colon \beta \in \mathbb{R}_{\geqslant 0}\}\,,\\
N_K(q_2) &= \operatorname{cone}\{-u\} = \{\beta(-u) \colon \beta \in \mathbb{R}_{\geqslant 0}\}\,.
\end{aligned}$$

for some suitable vector $u \in \mathbb{R}^2$. Let $v \neq 0$ be a vector orthogonal to u. We can move q_j along the edge F_j in at most two directions, namely v or $-v$. If we can move both q_1 and q_2 in the same direction, we simply translate the closed polygonal line q. If we can move q_1 and q_2 only in opposite directions, it is necessary to check whether the normal cones $N_T(p_1)$ and $N_T(p_2)$ allow such movement. If not, it is not possible to find suitable points $q_1 \in \operatorname{relint} F_1$, $q_2 \in \operatorname{relint} F_2$ (see Figure 4.7). This concludes our discussion of Algorithm 2.

Next, we discuss Algorithm 3. We choose edges F_1, F_2, F_3 of K. For each $j \in \{1,2,3\}$ we let u_j be the unique unit normal vector of K at some point in the relative interior of

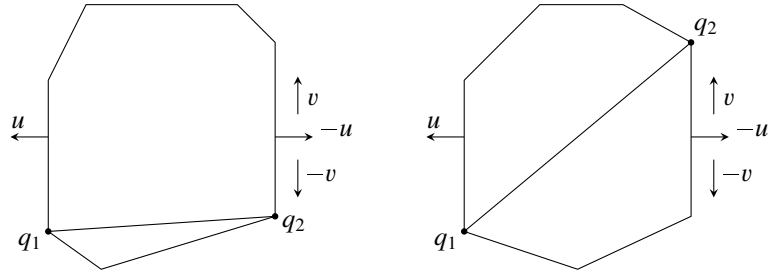


Figure 4.7: Two closed polygonal lines such that q_1, q_2 are vertices of K. On the left, we can translate q_1, q_2 upwards (in direction v). On the right, we need to move q_1 upwards and q_2 downwards. Whether this is possible depends on $N_T(p_1)$ and $N_T(p_2)$.

F_j. If possible we construct a triangle, i.e. a closed polygonal line with three vertices, γ by successively moving in the directions $-u_1, -u_2, -u_3$. We can do this easily by solving a system of linear equations:

$$0 = -u_1 - \mu_1 u_2 - \mu_2 u_3.$$

If this system does not have a unique, positive solution (μ_1, μ_2), we reject the iteration and proceed with the next choice of edges. If we find a unique, positive solution, then the task is to translate and scale γ with a positive factor such that its vertices lie on ∂T. This motivates the following definition.

Definition 4.3.7. *Let $\gamma \subseteq \mathbb{R}^2$ be a triangle and let $T \subseteq \mathbb{R}^2$ be a convex body. A (γ, T)-inbody is a set $\Delta \subseteq \mathbb{R}^2$ which fulfils:*

(i) $\Delta = \operatorname{conv}(\mu\gamma + b)$ *for some $\mu \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}^2$.*

(ii) All three vertices of Δ are contained in ∂T.

(iii) If $V(\Delta) = \{v_1, v_2, v_3\}$ denotes the set of vertices of Δ, then there is no line H through the origin, such that $N_T(v_1) \cup N_T(v_2) \cup N_T(v_3)$ is contained in one of the two closed halfspaces defined by H.

If we can find such a (γ, T)-inbody with vertices v_1, v_2, v_3, then we pick $p_j = v_j$ for every $j \in \{1, 2, 3\}$. The reason for the first two properties in Definition 4.3.7 are obvious: They ensure that we have

$$p_{j+1} - p_j = -\mu_{j+1} u_{j+1} \text{ for all } j \in \{1, 2, 3\},$$

where $p_4 := p_1$, $u_4 := u_1$ and $\mu_4 := \mu_1$. We take the convex hull in (i) to simplify some upcoming arguments. The reason for property (iii) is that we need to construct another triangle using directions $w_j \in N_T(v_j)$ for $j \in \{1, 2, 3\}$ later on in Algorithm 3. This is only possible if (iii) holds. We can find (γ, T)-inbodies with the following Lemma.

Lemma 4.3.8. *Let $T \subseteq \mathbb{R}^2$ be a polygon and let $\gamma \subseteq \mathbb{R}^2$ be a triangle. If there is a (γ, T)-inbody $\Delta^* = \operatorname{conv}(\mu^*\gamma + b^*)$, then (μ^*, b^*) is the unique solution of the following LP:*

$$\begin{aligned} \max \quad & \mu \\ \text{s. t.} \quad & \mu \geqslant 0,\ b \in \mathbb{R}^2, \\ & \mu\gamma + b \subseteq T. \end{aligned} \tag{4.9}$$

Proof. Let $\Delta = \operatorname{conv}\{v_1, v_2, v_3\}$ be a (γ, T)-inbody and let $\Delta^* = \operatorname{conv}\{v_1^*, v_2^*, v_3^*\}$ be as in the claim. Here, we choose the indices of the vertices such that there are $\mu > 0$ and $b \in \mathbb{R}^2$ with $v_j^* = \mu v_j + b$ for every $j \in \{1, 2, 3\}$. Note that such a labeling is possible since Δ is a homothetic copy of Δ^*. We start by letting H_1, H_2, H_3 be three lines defined by

$$v_2^*, v_3^* \in H_1\,,\ v_1^*, v_3^* \in H_2 \text{ and } v_1^*, v_2^* \in H_3.$$

Each of these lines is the affine hull of an edge of Δ^*. Furthermore, for every $j \in \{1, 2, 3\}$, H_j divides the plane $\mathbb{R}^2$ into two closed halfspaces H_j^+ and H_j^- such that $\Delta^* \subseteq H_j^+$. If $\Delta = \Delta^*$, then there is nothing to show. So, we assume $\Delta \neq \Delta^*$. Because Δ is a smaller or equal-sized homothetic copy of Δ^*, it is contained in H_j^+ for some $j \in \{1, 2, 3\}$. Without loss of generality, we assume $\Delta \subseteq H_1^+$, as we can treat the other cases similarly. This situation is depicted in Figure 4.8.

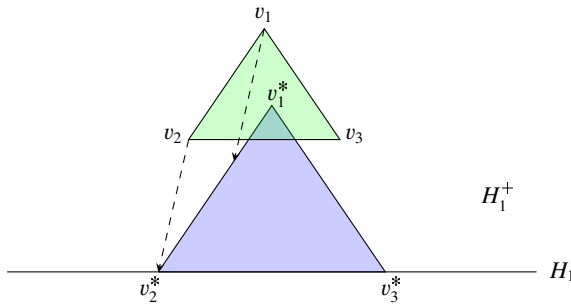


Figure 4.8: The (γ, T)-inbody Δ (green) with vertices v_1, v_2, v_3 and the triangle Δ^* (blue) with vertices v_1^*, v_2^*, v_3^*. The dashed arrows indicate the location of the line segment $[v_1, v_2]$ after shifting it by $v_2^* - v_2$.

We now show that property (iii) is violated for Δ. This contradiction implies $\Delta = \Delta^*$ and finishes the proof. More precisely, we claim that the halfspace

$$I = \{y \in \mathbb{R}^2 \colon (v_2^* - v_2)^T y \leqslant 0\}$$

contains $N_T(v_j)$ for every $j \in \{1, 2, 3\}$. By definition of the normal cone it immediately follows that $N_T(v_2) \subseteq I$. Because Δ is a smaller or equal-sized homothetic copy of Δ^*, we

have $v_2^* - v_3^* = \alpha(v_2 - v_3)$ for some $\alpha \geqslant 1$. For every $y \in N_T(v_3)$ this implies:

$$\begin{aligned}(v_2^* - v_2)^T y &= (v_2^* - v_2 + v_3^* - v_3^* + v_3 - v_3)^T y \\ &= \big((v_2^* - v_3^*) - (v_2 - v_3)\big)^T y + (v_3^* - v_3)^T y \\ &\leqslant (v_2^* - v_3^*)^T y - (v_2 - v_3)^T y \\ &= (\alpha - 1)(v_2 - v_3)^T y \\ &\leqslant 0.\end{aligned}$$

Thus, $N_T(v_3) \subseteq I$. Next, we observe that if we shift Δ by $v_2^* - v_2$, then the edge $[v_1, v_2]$ of Δ is contained in $[v_1^*, v_2^*]$ (see Figure 4.8). So, $v_1 + (v_2^* - v_2)$ is contained in $\Delta^* \subseteq T$. Now, for every $y \in N_T(v_1)$ we get:

$$0 \geqslant \big(v_1 + (v_2^* - v_2) - v_1\big)^T y = (v_2^* - v_2)^T y.$$

As desired this yields $N_T(v_1) \subseteq I$. □

We point out that there is not always a (γ, T)-inbody. For example, if (4.9) has multiple optimal solutions, then the proof shows that there is no (γ, T)-inbody. An example for this situation is depicted in Figure 4.9.

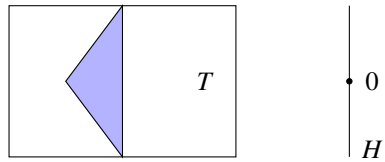


Figure 4.9: A situation where there is no (γ, T)-inbody. The blue area is a scaled translate of γ. Also shown is a line H through the origin 0.

As we can see there are multiple optimal solutions for (4.9) since we can shift the blue area to the left and right. The only way to have all vertices of this area on ∂T is to shift it to the left. Then all the corresponding normal vectors are contained in the halfspace on the left of H.

With Lemma 4.3.8 we can reduce the search of (γ, T)-inbodies to a maximization problem which we can formulate as an LP. It is clear that this problem has an optimal solution (μ^*, b^*) as long as T is compact. We check whether $\Delta^* = \mu^*\gamma + b^*$ fulfils properties (i) – (iii) in Definition 4.3.7. If at least one of these properties is violated, then there is no (γ, T)-inbody and we proceed with the next choice of F_1, F_2, F_3. Otherwise, we take a unit vector from $N_T(v)$ for each vertex v of Δ^*. We construct another triangle as before by successively moving in the directions of these vectors.

It is notable, that v may be a vertex of T. In this case $N_T(v)$ is not one-dimensional and the choice of the corresponding unit vector is not unique. One way to handle this case is due to Theorem 4.3.6: It is sufficient to find one vector in $N_T(v)$ such that the current iteration in Algorithm 3 yields a closed MBT $q = (q_1, q_2, q_3)$. If another unit vector in $N_T(v)$ yields a different closed MBT $q' = (q_1', q_2', q_3')$, then we have $q_j, q_j' \in F_j$ for every $j \in \{1, 2, 3\}$

and it follows from Theorem 4.3.6 that $\ell_T(q) = \ell_T(q')$. Therefore, we continue to pick unit vectors from $N_T(v)$ until we find one that yields a closed MBT. Alternatively, if we do not find such a unit vector in $N_T(v)$, we slightly perturb the vertices of K or T. Due to Lemma 2.3.3 we know that this changes the value of $c_{\mathrm{EHZ}}(K \times T)$ only slightly as well. The remainder of the algorithm is straight forward and only uses strategies that we discussed before.

Regarding efficiency, we point out that Algorithm 2 takes $O(|V(K)|^2 \cdot |V(T)|^2)$ iterations before it terminates. This is clear since the number of edges of a polygon equals the number of its vertices. For each of the two polygons K and T the algorithm considers at most one choice for F_1, F_2, G_1, G_2 per iteration. In each iteration we search for the points q_1, q_2, p_1, p_2. In the worst case (i.e. if F_1, F_2, G_1, G_2 are edges) we solve an LP with 10 variables and $2(|V(K)| + |V(T)| + 3)$ constraints. We use the `conelp` solver of CVXOPT [97], which relies on a primal-dual path-following method.

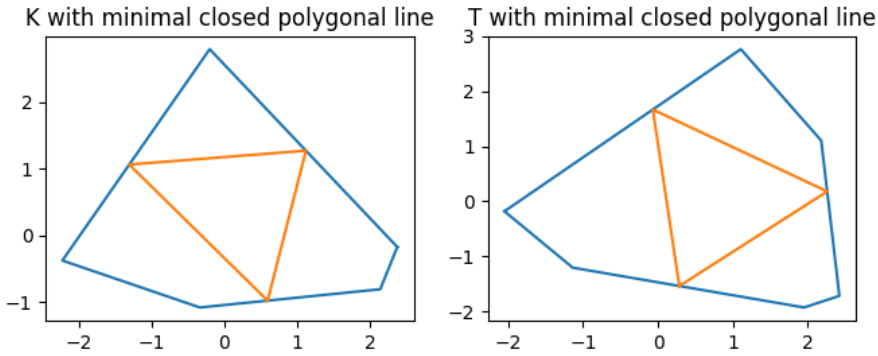


Figure 4.10: An exemplary output of Algorithm 3. The randomly generated polygons (K on the left and T on the right) are depicted in blue. The calculated closed polygonal lines (q on the left and p on the right) are depicted in orange. The closed polygonal line q in the picture has smaller ℓ_T-length than the closed polygonal line with two vertices that we find with Algorithm 2.

Algorithm 3 takes $O(|V(K)|^3)$ iterations to consider every choice of edges F_1, F_2, F_3. In each iteration we solve the maximization problem stated in Lemma 4.3.8. This is an LP with three variables and $3|V(T)| + 1$ constraints. We can implement the remainder of the loop for F_1, F_2, F_3 with running time $O(|V(K)|)$. Finally, we note that the calculations for each choice of edges are independent of each other. Therefore, we use parallel computing to speed up the calculations.

In Table 6.2 in the appendix we examine the running time of our implementation of Algorithm 2 and 3. We let K and T be polygons and consider three different cases. First, we regard the case where both K and T have the same number of vertices. In the second case K has a small number of vertices and in the third case we choose T to have few vertices. Every time we choose the polygons K and T randomly in the following sense. First, we generate some normally distributed points in $\mathbb{R}^2$ and scale each of these points to have a random length in $[1, 3]$. Then we get a polygon by taking the convex hull of these

random points. As the number of random points grows, the convex hull resembles a ball of radius 3 due to the normal distribution. So, to accelerate this process, we reduce the interval $[1,3]$ for polytopes with many vertices ($\geqslant 30$).

We compare the running time for finding a closed polygonal line with minimal ℓ_T-length and with two vertices to the running time for finding one with three vertices. As we can see the running time for two vertices is approximately symmetric in $\big|V(K)\big|$ and $\big|V(T)\big|$. In contrast to this, the running time for three vertices mainly depends on $\big|V(K)\big|$.

4.3.4 The Minkowski setting in higher dimension

To conclude Chapter 4 we point out a slightly different approach to compute a closed MBT with minimal ℓ_T-length that is also applicable if the polytopes K and T have dimension higher than two. Instead of constructing, scaling and translating closed polygonal lines, we attempt to find a solution of (4.1) immediately. Assume we search for a closed MBT with minimal ℓ_T-length and m bouncing points, where $m \in \{2,\dots,n+1\}$ and n is the dimension of K and T. Then we fix faces $F_1,\dots,F_m$ of K and faces $G_1,\dots,G_m$ of T and solve the following problem:

$$\begin{aligned}&\text{Find } q_1,\dots,q_m,p_1,\dots,p_m \text{ such that:}\\ &\quad q_j \in \operatorname{relint} F_j,\ p_j \in \operatorname{relint} G_j \ \text{ for all } j\in\{1,\dots,m\},\\ &\quad q_{j+1}-q_j \in N_T(p_j) \ \text{ for all } j\in\{1,\dots,m\},\\ &\quad p_{j+1}-p_j \in -N_K(q_{j+1}) \ \text{ for all } j\in\{1,\dots,m\}.\end{aligned} \tag{4.10}$$

Here, $q_{m+1} := q_1$ and $p_{m+1} := p_1$. If $m = n+1$ it is sufficient to consider facets $F_1,\dots,F_m$ instead of faces. Note that if we fix a face G_j of T, we only have to compute the one-dimensional normal cones at facets of T that contain G_j. If $w_1,\dots,w_k$ are the unique unit vectors in these normal cones, we have

$$N_T(p_j) = \operatorname{cone}\{w_1,\dots,w_k\}$$

for every $p_j \in \operatorname{relint} G_j$. Thus, we can ask whether $q_{j+1}-q_j$ lies in $N_T(p_j)$ by asking whether there are nonnegative numbers $\lambda_1,\dots,\lambda_k$ such that

$$q_{j+1}-q_j = \sum_{i=1}^{k} \lambda_i w_i.$$

We can rephrase all constraints in (4.10) that contain a normal cone to get an LP. If we solve this LP for every choice of faces $F_1,\dots,F_m$ and $G_1,\dots,G_m$ and for every $m \in \{2,\dots,n+1\}$, we find a closed MBT with minimal ℓ_T-length. To accelerate this process we can also divide (4.10) into two problems. First, we solve the following problem:

$$\begin{aligned}&\text{Find } p_1,\dots,p_m \text{ such that:}\\ &\quad p_j \in \operatorname{relint} G_j \ \text{ for all } j\in\{1,\dots,m\},\\ &\quad p_{j+1}-p_j \in -N_K(q_{j+1}) \ \text{ for all } j\in\{1,\dots,m\},\end{aligned}$$

with some $q_j \in \operatorname{relint} F_j$ for every $j \in \{1,\dots,m\}$. Here, $p_{m+1} = p_1$ and $q_{m+1} = q_1$. If we do not find a solution, then we can reject the iteration and proceed with the next

choice of faces. Otherwise, we solve a similar problem to determine $q_1, \dots, q_m$. In the two-dimensional case this approach is much slower than running Algorithm 2 and 3. The reason is that we now iterate over every choice of F_1, F_2, F_3 and G_1, G_2, G_3. In contrast to this, Algorithm 3 only iterates over every choice of F_1, F_2, F_3 because Lemma 4.3.8 provides us with an easy way to find (p_1, p_2, p_3).

Chapter Five

Bounds on the EHZ capacity of polytopes

In this chapter we examine a formulation of the EHZ capacity that is due to Haim-Kislev [47] (see identity (5.4) below). This formulation is similar to the one that is given by Theorem 2.3.6, in the sense that they both express $1/4c_{\mathrm{EHZ}}(C)$, where C is a convex set, as a maximization problem. The difference is that in Theorem 2.3.6 we require C to have smooth boundary, while Haim-Kislev considers polytopes instead. In Chapter 5.1 we compare these two expressions in more detail. Afterwards, in Chapter 5.2 and 5.3 we proceed to compute upper bounds on the EHZ capacity of polytopes using Haim-Kislev's formulation. To this end, we write (5.4) as a QAP and adapt the technique that is mentioned in Chapter 3 to bound the optimal value. Another strategy that we pursue is to reformulate (5.4) using multiple CPs. Then, as noted in Chapter 3, we employ semidefinite relaxation to compute upper bounds. In Chapter 5.4 we compute lower bounds on the optimal value as well.

5.1 The EHZ capacity as a maximization problem

We recall Theorem 2.3.6. If $C \subseteq \mathbb{R}^{2n}$ is a convex set with smooth boundary and $0 \in \operatorname{int} C$, then:

$$\frac{1}{4c_{\mathrm{EHZ}}(C)} = \max \left\{ \mathbb{A}(x) \mid x\colon [0,1] \to \mathbb{R}^{2n} \text{ absolutely continuous}, \; x(0) = x(1), \;\; \dot{x} \in C^\circ \text{ a.e.} \right\}. \tag{5.1}$$

We observe that finding an optimal solution is difficult because the set of feasible solutions has infinite dimension. Therefore, we would like to reduce the search to a finite-dimensional subset that contains optimal solutions. A common way to find such a subset is to discretize the original problem. More precisely, we add the constraint that x is a closed polygonal line. In other words, we assume there are $0 = t_0 \leqslant \ldots \leqslant t_k = 1$ and $v_1, \ldots, v_k \in C^\circ$ for some integer $k \geqslant 1$ such that for every $i \in \{1, \ldots, k\}$:

$$\dot{x}(t) = v_i \quad \text{for all } t \in (t_{i-1}, t_i).$$

In particular, x is linear on the interval $[t_{i-1}, t_i]$ for every $i \in \{1, \dots, k\}$. Thus, we have

$$x(t) = x(t_{i-1}) + (t - t_{i-1})v_i \quad \forall\, t \in [t_{i-1}, t_i],\ i \in \{1, \dots, k\}.$$

It follows that if $t \in [t_{i-1}, t_i]$, then we can write

$$\begin{aligned} x(t) = (t - t_{i-1})v_i + x(t_{i-1}) = (t - t_{i-1})v_i + (t_{i-1} - t_{i-2})v_{i-1} + x(t_{i-2}) = \dots \\ = (t - t_{i-1})v_i + \left(\sum_{j=1}^{i-1}(t_j - t_{j-1})v_j\right) + x(t_0). \end{aligned}$$

In (5.1) we require that $x(0) = x(1)$ holds. Therefore:

$$0 = x(t_k) - x(t_0) = \left(\sum_{i=1}^{k}(t_i - t_{i-1})v_i\right) + x(t_0) - x(t_0) = \sum_{i=1}^{k} y_i v_i,$$

where we substitute $y_i := t_i - t_{i-1}$. The objective function takes the following form:

$$\begin{aligned} \mathbb{A}(x) &= -\frac{1}{2}\int_0^1 \dot{x}(t)^T J x(t)\,dt \\ &= -\frac{1}{2}\sum_{i=1}^{k}\int_{t_{i-1}}^{t_i} v_i^T J\left[\left(\sum_{j=1}^{i-1}(t_j - t_{j-1})v_j\right) + (t - t_{i-1})v_i\right] dt \\ &= -\frac{1}{2}\sum_{i=1}^{k}\left[\int_{t_{i-1}}^{t_i}\left(\sum_{j=1}^{i-1}(t_j - t_{j-1})v_i^T J v_j\right) dt + \int_{t_{i-1}}^{t_i}(t - t_{i-1})v_i^T J v_i\,dt\right] \\ &= -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1}(t_i - t_{i-1})(t_j - t_{j-1})v_i^T J v_j \\ &= -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} y_i y_j v_i^T J v_j, \end{aligned}$$

where we use the fact that $v^T J v = 0$ for every $v \in \mathbb{R}^{2n}$ because J is skew-symmetric. Two additional constraints arise from the discretization. First, we require that $t_{i-1} \leqslant t_i$ or equivalently $y_i \geqslant 0$ for every $i \in \{1, \dots, k\}$. Second, the numbers $t_0, \dots, t_k$ discretize an interval of length one. Thus,

$$t_k - t_0 = 1 \iff \sum_{i=1}^{k}(t_i - t_{i-1}) = 1 \iff \sum_{i=1}^{k} y_i = 1.$$

Altogether, we receive the following optimization problem as a discretization of (5.1). Note that $k \geqslant 1$ is some fixed integer and that the optimization variables are a vector $y \in \mathbb{R}^k$ and

points $v_1, \dots, v_k \in C^\circ$.

$$\begin{aligned} \max \quad & -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} y_i y_j v_i^T J v_j \\ \text{s. t.} \quad & \sum_{i=1}^{k} y_i v_i = 0 \\ & \sum_{i=1}^{k} y_i = 1 \\ & v_1, \dots, v_k \in C^\circ, \;\; y \in \mathbb{R}^k_{\geqslant 0}. \end{aligned} \tag{5.2}$$

One could assume that the optimum is only a lower bound on $1/4c_{\mathrm{EHZ}}(C)$ since we added the constraint that x is a closed polygonal line. However, Haim-Kislev [47] showed that if C is a polytope, if we pick $k = |V(C^\circ)|$ and if we require that $v_1, \dots, v_k$ are exactly the vertices of C°, then this maximum is equal to $1/4c_{\mathrm{EHZ}}(C)$.

Theorem 5.1.1. *[47] Let $C \subseteq \mathbb{R}^{2n}$ be a polytope with $0 \in \operatorname{int} C$ and let k be the number of facets of C. Let $A \in \mathbb{R}^{k\times 2n}$ and $b \in \mathbb{R}^k$ such that*

$$C = \{z \in \mathbb{R}^{2n} \colon Ax \leqslant b\}$$

and such that the ith row a_i^T of A satisfies $||a_i|| = 1$ for every $i \in \{1, \dots, k\}$. Then:

$$\begin{aligned} \frac{1}{4c_{\mathrm{EHZ}}(C)} = \quad \max \quad & -\sum_{i=1}^{k}\sum_{j=1}^{i-1} \beta_{\sigma(i)}\beta_{\sigma(j)} a_{\sigma(i)}^T J a_{\sigma(j)} \\ \text{s. t.} \quad & \sum_{i=1}^{k} \beta_i a_i = 0 \\ & \sum_{i=1}^{k} \beta_i b_i = 1 \\ & \beta_i \geqslant 0, \; \forall\, i \in \{1, \dots, k\} \\ & \sigma \in \mathrm{Sym}_k. \end{aligned}$$

To observe the connection between (5.2) and the maximization problem in Theorem 5.1.1 we first characterize the vertices of C°. This characterization is well known and we state the proof from [105] for the sake of completeness.

Theorem 5.1.2. *[105] Let $C \subseteq \mathbb{R}^n$ be a polytope with $0 \in \operatorname{int} C$. Furthermore, assume that*

$$C = \{z \in \mathbb{R}^n \colon Az \leqslant e\}$$

for some matrix $A \in \mathbb{R}^{m\times n}$, where $e = (1, \dots, 1)^T \in \mathbb{R}^m$. Then

$$C^\circ = \operatorname{conv}\{a_i \colon i \in \{1, \dots, m\}\},$$

where a_i^T denotes the ith row of A for $i \in \{1, \dots, m\}$.

Proof. Let $\lambda_1, \dots, \lambda_m \geqslant 0$ such that $\sum_i \lambda_i = 1$. Then for every $x \in C$ we get

$$\left(\sum_{i=1}^{m} \lambda_i a_i\right)^T x = \sum_{i=1}^{m} \lambda_i a_i^T x \leqslant \sum_{i=1}^{m} \lambda_i = 1.$$

Therefore, every convex combination of $a_1, \dots, a_m$ is contained in C°.

For the converse, let $z \in C^\circ$. By definition, $z^T c \leqslant 1$ holds for every $c \in C$. A variant of Farkas' lemma (see [105]) now states that there is a vector $\lambda \in \mathbb{R}^m_{\geqslant 0}$ such that $\lambda^T A = z^T$ and $\lambda^T e \leqslant 1$, where $e \in \mathbb{R}^m$ is the vector whose entries are all equal to one. Next, we note that $Au \leqslant -e$ has no solution $u \in \mathbb{R}^m$. Otherwise, if $Au \leqslant -e$ holds for some u, then we have $u \neq 0$ and $\alpha Au \leqslant -e \leqslant e$ for every $\alpha \in \mathbb{R}_{\geqslant 0}$. In particular, we find that $\alpha u \in C$ for every $\alpha \geqslant 0$, which contradicts the fact that C is bounded. Now, since $Au \leqslant -e$ has no solution, Farkas' lemma (see [105]) implies that there is a vector $\mu \in \mathbb{R}^m_{\geqslant 0}$ such that $\mu^T A = 0$ and $\mu^T e > 0$. We let

$$\nu = \lambda + \frac{1 - \lambda^T e}{\mu^T e}\mu$$

and observe that $\nu \geqslant 0$ because $\lambda, \mu \geqslant 0$ and $\lambda^T e \leqslant 1$. Furthermore, we have

$$\nu^T e = \lambda^T e + \frac{1 - \lambda^T e}{\mu^T e}\mu^T e = 1,$$
$$\nu^T A = \underbrace{\lambda^T A}_{=z^T} + \frac{1 - \lambda^T e}{\mu^T e} \underbrace{\mu^T A}_{=0} = z^T.$$

Hence, z is a convex combination of the rows of A. □

Let C, A, b as in Theorem 5.1.1. With Theorem 5.1.2 we get

$$C^\circ = \operatorname{conv}\left\{\frac{a_i}{b_i} : i \in \{1, \dots k\}\right\}. \tag{5.3}$$

If we require that $v_1, \dots, v_k$ are the vertices of C° in (5.2), then we can introduce the optimization variable $\sigma \in \mathrm{Sym}_k$ and let $v_i = a_{\sigma(i)}/b_{\sigma(i)}$ for every $i \in \{1, \dots, k\}$. The objective function in (5.2) becomes:

$$-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} y_i y_j v_i^T J v_j = -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} \frac{y_i}{b_{\sigma(i)}}\frac{y_j}{b_{\sigma(j)}} a_{\sigma(i)}^T J a_{\sigma(j)}.$$

So, we get the same objective function as in Theorem 5.1.1 if we let $\beta_{\sigma(i)} = y_i/b_{\sigma(i)}$ for every $i \in \{1, \dots, k\}$. Furthermore, we get

$$\sum_{i=1}^{k} y_i v_i = 0 \iff \sum_{i=1}^{k} \beta_{\sigma(i)} b_{\sigma(i)} \frac{a_{\sigma(i)}}{b_{\sigma(i)}} = 0 \iff \sum_{i=1}^{k} \beta_i a_i = 0$$

as well as

$$\sum_{i=1}^{k} y_i = 1 \iff \sum_{i=1}^{k} \beta_{\sigma(i)} b_{\sigma(i)} = 1 \iff \sum_{i=1}^{k} \beta_i b_i = 1.$$

Since we require that $0 \in \operatorname{int} C$, the numbers $b_1, \ldots, b_k$ are positive. So, we have that $y_1, \ldots, y_k \geqslant 0$ if and only if $\beta_1, \ldots, \beta_k \geqslant 0$.

The goal in the remainder of this chapter is to provide upper and lower bounds on $1/4c_{\mathrm{EHZ}}(C)$ and consequently also on $c_{\mathrm{EHZ}}(C)$ by investigating the maximization problem

$$\begin{aligned} \frac{1}{4c_{\mathrm{EHZ}}(C)} = \max \quad & -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{i-1} y_i y_j v_{\sigma(i)}^T J v_{\sigma(j)} \\ \text{s. t.} \quad & \sum_{i=1}^{k} y_i v_{\sigma(i)} = 0 \\ & \sum_{i=1}^{k} y_i = 1 \\ & y_i \geqslant 0 \ \ \forall\, i \in \{1, \ldots, k\} \\ & \sigma \in \mathrm{Sym}_k, \end{aligned} \tag{5.4}$$

where $v_1, \ldots, v_k$ are the vertices of C°. For this, C always denotes a polytope with $0 \in \operatorname{int} C$ and k denotes the number of facets of C or equivalently the number of vertices of C°.

5.2 An upper bound via QAP techniques

To start our investigation of (5.4) we focus on the fact that we ask for an optimal permutation $\sigma \in \mathrm{Sym}_k$. Thus, (5.4) is reminiscent of a QAP. More precisely, if we substitute $z_{\sigma(j)} := y_j$ for every $j \in \{1, \ldots, k\}$, we get

$$\begin{aligned} \frac{1}{4c_{\mathrm{EHZ}}(C)} &= \max \left\{ -\frac{1}{2} \sum_{h=1}^{k} \sum_{j=1}^{h-1} z_{\sigma(h)} z_{\sigma(j)} v_{\sigma(h)}^T J v_{\sigma(j)} : z \in M,\ \sigma \in \mathrm{Sym}_k \right\}, \\ M &= \left\{ z \in \mathbb{R}^k_{\geqslant 0} \colon \sum_{j=1}^{k} z_j v_j = 0,\ \sum_{j=1}^{k} z_j = 1 \right\}. \end{aligned}$$

Note that the set M is the $(k-1)$-dimensional standard simplex intersected with the kernel of the matrix $(v_1, \ldots, v_k)$. Furthermore, M does not depend on σ. Thus, we can write the maximum as two maxima:

$$\frac{1}{4c_{\mathrm{EHZ}}(C)} = \frac{1}{2} \max_{z \in M} \left(\max_{\sigma \in \mathrm{Sym}_k} -\sum_{h=1}^{k} \sum_{j=1}^{h-1} z_{\sigma(h)} z_{\sigma(j)} v_{\sigma(h)}^T J v_{\sigma(j)} \right). \tag{5.5}$$

Recall from Chapter 3 that a $(k \times k)$-matrix X is called a *permutation matrix* if there is a permutation $\sigma \in \mathrm{Sym}_k$ such that

$$X_{h,j} = \begin{cases} 1, \text{ if } h = \sigma(j), \\ 0, \text{ otherwise.} \end{cases}$$

Furthermore, we denote the set of all $(k \times k)$-permutation matrices with Π_k. We observe that the maximization over Sym_k is a QAP of the form

$$\max_{X \in \Pi_k} \left(\left\langle F, X^T D(z) X \right\rangle + \langle B, X \rangle \right), \tag{5.6}$$

where we define the matrices $F, D(z), B \in \mathbb{R}^{k \times k}$ by

$$B_{h,j} = 0, \quad D(z)_{h,j} = -z_h z_j v_h^T J v_j, \quad F_{h,j} = \begin{cases} 1, \text{ if } h > j, \\ 0, \text{ if } h \leqslant j, \end{cases}$$

for all $h, j \in \{1, \ldots, k\}$. Here, it is notable that the matrix $D(z)$ depends on some vector $z \in M$. We point out that (5.6) has the same form as (3.1) except for the fact that we perform a maximization instead of a minimization. This is not an issue since the approach in Chapter 3.1 still suggests that we give an upper and a lower bound on the objective function in (5.6) that only depends on the eigenvalues of F and $D(z)$. In particular, we do this regardless of whether the goal is to maximize or minimize this objective function. A more pressing problem is that the reasoning in Chapter 3.1 mainly relies on Theorem 3.1.1, which requires the matrices F and $D(z)$ to be symmetric. This is not the case here. Due to the skew-symmetry of the matrix J it is obvious that $D(z)$ is skew-symmetric. F is neither symmetric nor skew-symmetric. However, we can consider $F' := (F - F^T)/2$ instead. Then F' is skew-symmetric and

$$\begin{aligned} \left\langle F', X^T D(z) X \right\rangle &= \frac{1}{2} \left(\left\langle F, X^T D(z) X \right\rangle - \left\langle F^T, X^T D(z) X \right\rangle \right) \\ &= \frac{1}{2} \left(\left\langle F, X^T D(z) X \right\rangle + \left\langle F^T, X^T (-D(z)) X \right\rangle \right) \\ &= \frac{1}{2} \left(\left\langle F, X^T D(z) X \right\rangle + \left\langle F^T, (X^T D(z) X)^T \right\rangle \right) = \left\langle F, X^T D(z) X \right\rangle. \end{aligned} \tag{5.7}$$

The goal is now to prove another version of Theorem 3.1.1 for skew-symmetric matrices. First, we recall some properties of skew-symmetric matrices and provide a proof for the sake of completeness.

Lemma 5.2.1. *[54] Let $A \in \mathbb{R}^{k \times k}$ be a real, skew-symmetric matrix. Then:*

(i) All eigenvalues of A are purely imaginary.

(ii) If λ is an eigenvalue of A, then so is $-\lambda$.

(iii) A is unitarily diagonalizable, i.e. there is some unitary matrix $U \in \mathbb{C}^{k \times k}$ such that $A = U \Sigma U^*$, where Σ is a diagonal matrix containing all eigenvalues of A as diagonal elements. Furthermore, the columns of U are exactly the eigenvectors of A.

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue for the eigenvector $x \neq 0$. Then we have

$$\overline{\lambda} x^* x = \overline{\lambda x^T} x = \overline{(Ax)^T} x = x^* A^* x = -x^* A x = -\lambda x^* x.$$

It follows that $\overline{\lambda} = -\lambda$ which implies that $\operatorname{Re} \lambda = 0$. Furthermore,

$$Ax = \lambda x \implies -Ax = -\lambda x \implies A^T x = -\lambda x.$$

Hence, $-\lambda$ is an eigenvalue of A^T. Because A and A^T have the same eigenvalues, statement (ii) follows.

For (iii) we note that A is a normal matrix, i.e. $AA^* = A^*A$, because we have $A^* = A^T = -A$. We consider a Schur decomposition of A:

$$A = URU^*,$$

where $R \in \mathbb{C}^{k \times k}$ is an upper triangular matrix and $U \in \mathbb{C}^{k \times k}$ is a unitary matrix. The fact that A is normal implies that R is normal as well:

$$RR^* = U^*AUU^*A^*U = U^*AA^*U = U^*A^*AU = U^*A^*UU^*AU = R^*R.$$

It remains to show that a normal upper triangular $(k \times k)$-matrix is a diagonal matrix. We prove this inductively. If $k = 1$ it is obvious that R is a diagonal matrix. If $k > 1$ we write

$$R = \begin{pmatrix} R_{1,1} & v^* \\ 0 & R' \end{pmatrix},$$

where $v \in \mathbb{C}^{k-1}$ and $R' \in \mathbb{C}^{(k-1)\times(k-1)}$ is an upper triangular matrix. With this, we get:

$$\begin{pmatrix} |R_{1,1}|^2 + ||v||^2 & v^*(R')^* \\ R'v & R'(R')^* \end{pmatrix} = RR^* = R^*R = \begin{pmatrix} |R_{1,1}|^2 & \overline{R_{1,1}} v^* \\ R_{1,1}v & vv^* + (R')^*R' \end{pmatrix}. \tag{5.8}$$

In particular, $|R_{1,1}|^2 + ||v||^2 = |R_{1,1}|^2$ implies that $v = 0$. If we plug this into (5.8) we also get $R'(R')^* = (R')^*R'$. So, R has the form

$$R = \begin{pmatrix} R_{1,1} & 0 \\ 0 & R' \end{pmatrix}$$

with a normal upper triangular $(k-1)$ by $(k-1)$ matrix R'. By induction it follows that R is a diagonal matrix. Moreover, the diagonal entries of R are the eigenvalues of A because A and R are similar matrices. It is straightforward to show that the columns of U are the eigenvectors of A. Let u_j be the jth column for $j \in \{1, \ldots, k\}$. If $e_j \in \mathbb{R}^k$ denotes the jth unit vector, it follows that

$$Au_j = U\Sigma U^* u_j = U\Sigma e_j = U\Sigma_{j,j} e_j = \Sigma_{j,j} u_j. \qquad \square$$

We now pursue a similar strategy as in [26] to prove a suitable version of Theorem 3.1.1. To this end, we recall from Chapter 3 that a matrix with nonnegative entries is called *doubly stochastic* if all its row and column sums are equal to one.

Theorem 5.2.2. *Let $Y_1, Y_2 \in \mathbb{R}^{k\times k}$ be skew-symmetric matrices. Let $\lambda \in \mathbb{R}^k$ be the vector of eigenvalues of Y_1 and let $x_1, \dots, x_k$ be the corresponding orthonormal eigenvectors. Similarly, let $\mu \in \mathbb{R}^k$ be the vector of eigenvalues of Y_2 and let $y_1, \dots, y_k$ be the corresponding orthonormal eigenvectors. Then:*

$$(i) \qquad \langle Y_1, Y_2\rangle = \lambda^T S \mu, \quad \text{where } S = \left(|x_h^T y_j|^2\right)_{h,j\in\{1,\dots,k\}} \text{ is a doubly stochastic matrix.}$$

$$(ii) \qquad \langle Y_1, Y_2\rangle \leqslant -\min_{\sigma\in \mathrm{Sym}_k} \sum_{j=1}^{k} \mathrm{Im}\,\lambda_j \cdot \mathrm{Im}\,\mu_{\sigma(j)}.$$

Theorem 5.2.2 provides us with an upper bound on $\left\langle F', X^T D(z) X\right\rangle$ that only depends on the eigenvalues of F' and $D(z)$ since $D(z)$ and $X^T D(z)X$ have the same eigenvalues. We point out that a similar situation is considered by Hoffman and Wielandt [52]. They show that there is a permutation $\sigma \in \mathrm{Sym}_k$ such that

$$\sum_{j=1}^{k} |\lambda_{\sigma(j)} - \mu_{\sigma(j)}|^2 \leqslant \langle A - B, A - B\rangle, \tag{5.9}$$

where $A, B \in \mathbb{R}^{k\times k}$ are normal matrices, i.e. $AA^* = A^*A$ and $BB^* = B^*B$, with eigenvalues $\lambda_1, \dots, \lambda_k$ and $\mu_1, \dots, \mu_k$. The proof of Theorem 5.2.2, that we present here, is similar to the proof of (5.9) by Hoffman and Wielandt. Additionally, it is noted in [52] that finding the permutation σ in (5.9) is difficult in general and that it is easy to find σ in the case that A and B are Hermitian. Later, we make a similar observation for the skew-symmetric case, i.e. we deduce a permutation for which the minimum in Theorem 5.2.2 (ii) is attained.

Given the properties of skew-symmetric matrices in Lemma 5.2.1, the proof of (i) is a simple calculation. In addition, we apply the following lemma to show that the matrix S in Theorem 5.2.2 is indeed doubly stochastic.

Lemma 5.2.3. *Let $\{x_1, \dots, x_k\}$ and $\{y_1, \dots, y_k\}$ be orthonormal bases of $\mathbb{C}^k$. The matrix $S \in \mathbb{R}^{k\times k}$ given by*

$$S_{h,j} = |x_h^T y_j|^2$$

for every $h, j \in \{1, \dots, k\}$ is doubly stochastic.

Proof. We need to show that every entry of S is nonnegative and that every row and column sum is equal to one. The nonnegativity of the entries is obvious from the definition of S.

We fix an index $j \in \{1, \dots, k\}$. Since $\{x_1, \dots, x_k\}$ is an orthonormal basis of $\mathbb{C}^k$, $\{\bar{x_1}, \dots, \bar{x_k}\}$ is an orthonormal basis of $\mathbb{C}^k$ as well. So, we can express y_j in terms of these basis elements, i.e.:

$$y_j = \sum_{\ell} \alpha_\ell \bar{x_\ell},$$

for some $\alpha_1, \dots, \alpha_\ell \in \mathbb{C}$. Thus, the jth column sum is

$$\sum_{h=1}^{k} S_{h,j} = \sum_{h=1}^{k} |x_h^T y_j|^2 = \sum_{h=1}^{k} \left|\sum_{\ell=1}^{k} \alpha_\ell x_h^T \bar{x_\ell}\right|^2 = \sum_{h=1}^{k} \left|\sum_{\ell=1}^{k} \alpha_\ell \bar{x_h}^* \bar{x_\ell}\right|^2 = \sum_{h=1}^{k} |\alpha_h|^2.$$

To show that this sum is equal to one, we use that $\{y_1, \ldots, y_k\}$ is a basis of $\mathbb{C}^k$ as well:

$$1 = y_j^* y_j = \left(\sum_{\ell=1}^{k} \alpha_\ell \bar{x}_\ell\right)^* \left(\sum_{r=1}^{k} \alpha_r \bar{x}_r\right) = \sum_{\ell=1}^{k}\sum_{r=1}^{k} \bar{\alpha}_\ell \alpha_r \bar{x}_\ell{}^* \bar{x}_r$$
$$= \sum_{\ell=1}^{k} \alpha_\ell \bar{\alpha}_\ell = \sum_{\ell=1}^{k} |\alpha_\ell|^2.$$

If we fix a row index instead, then a similar calculation also yields that the corresponding row sum of S is equal to one. □

The second statement in Theorem 5.2.2 follows from the first one if we apply the famous Birkhoff-von Neumann theorem [19].

Theorem 5.2.4. *[19] The set of all doubly stochastic $(k \times k)$-matrices is a polytope and its vertex set is Π_k.*

In other words, a matrix is doubly stochastic if and only if it is a convex combination of permutation matrices. We now give the complete proof of Theorem 5.2.2.

Proof of Theorem 5.2.2. We start with the proof of statement (i). According to Lemma 5.2.1 we can write

$$Y_1 = \sum_{j=1}^{k} \lambda_j x_j x_j^* \quad \text{and} \quad Y_2 = \sum_{j=1}^{k} \mu_j y_j y_j^*.$$

Thus, we have

$$\langle Y_1, Y_2 \rangle = \sum_{h=1}^{k}\sum_{j=1}^{k} \lambda_h \mu_j \left\langle x_h x_h^*, y_j y_j^* \right\rangle = \sum_{h=1}^{k}\sum_{j=1}^{k} \lambda_h \mu_j \operatorname{tr}(\bar{x}_h x_h^T y_j \bar{y}_j{}^T)$$
$$= \sum_{h=1}^{k}\sum_{j=1}^{k} \lambda_h \mu_j \operatorname{tr}(x_h^T y_j \bar{y}_j{}^T \bar{x}_h) = \sum_{h=1}^{k}\sum_{j=1}^{k} \lambda_h \mu_j |x_h^T y_j|^2 = \lambda^T S \mu,$$

where the matrix $S \in \mathbb{R}^{k\times k}$ is defined by $S_{h,j} = |x_h^T y_j|^2$ for every $h, j \in \{1, \ldots, k\}$. Moreover, S is doubly stochastic due to Lemma 5.2.3.

To show (ii) we apply the Birkhoff-von Neumann theorem and find that S is a convex combination of permutation matrices. This means that there is a number $\alpha_X \geqslant 0$ for every $X \in \Pi_k$ such that

$$S = \sum_{X\in\Pi_k} \alpha_X X \quad \text{and} \quad \sum_{X\in\Pi_k} \alpha_X = 1.$$

In total, we get:

$$\begin{aligned}\langle Y_1, Y_2\rangle = \lambda^T S\mu &= \sum_{X\in\Pi_k} \alpha_X \lambda^T X\mu = -\sum_{X\in\Pi_k} \alpha_X \,(\operatorname{Im}\lambda)^T\, X\,(\operatorname{Im}\mu)\\ &\leqslant -\sum_{X\in\Pi_k} \alpha_X \min_{Y\in\Pi_k} (\operatorname{Im}\lambda)^T\, Y\,(\operatorname{Im}\mu) \qquad (5.10)\\ &= -\min_{\sigma\in\mathrm{Sym}_k} \sum_{j=1}^{k} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_{\sigma(j)}.\end{aligned}$$

Here, we use Lemma 5.2.1 (i) in the first line of equation (5.10). □

As mentioned earlier, Theorem 5.2.2 provides an upper bound on $\left\langle F', X^T D(z) X\right\rangle$ for every $X \in \Pi_k$. Moreover, this bound depends only on the eigenvalues of F' and $D(z)$. In particular, we have

$$\max_{X\in\Pi_k} \left\langle F', X^T D(z) X\right\rangle \leqslant -\min_{\sigma\in\mathrm{Sym}_k} \sum_{j=1}^{k} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_{\sigma(j)}, \qquad (5.11)$$

where $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of F' and $\mu_1, \ldots, \mu_k$ are the eigenvalues of $D(z)$. Now, we compute this upper bound. Let us order the eigenvalues of F' and $D(z)$ in the following way:

$$\begin{aligned}\operatorname{Im}\lambda_1 &\geqslant \operatorname{Im}\lambda_2 \geqslant \ldots \geqslant \operatorname{Im}\lambda_k,\\ \operatorname{Im}\mu_1 &\geqslant \operatorname{Im}\mu_2 \geqslant \ldots \geqslant \operatorname{Im}\mu_k.\end{aligned}$$

We can use Lemma 5.2.1 (ii) to pick up where we left off in (5.11):

$$\begin{aligned}-\min_{\sigma\in\mathrm{Sym}_k} \sum_{j=1}^{k} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_{\sigma(j)} &= -\sum_{j=1}^{k} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_{k-j+1}\\ &= -2\sum_{j=1}^{\lfloor \frac{k}{2}\rfloor} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_{k-j+1} \qquad (5.12)\\ &= 2\sum_{j=1}^{\lfloor \frac{k}{2}\rfloor} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_j.\end{aligned}$$

For the second equality in (5.12) we point out that the eigenvalues $\lambda_{\lfloor k/2\rfloor+1}$ and $\mu_{\lfloor k/2\rfloor+1}$ are zero if k is odd. This is due to Lemma 5.2.1 (ii).

It remains to find the eigenvalues $\lambda_1, \ldots, \lambda_k$ of F' and $\mu_1, \ldots, \mu_k$ of $D(z)$. Finding $\lambda_1, \ldots, \lambda_k$ is easy since we can calculate them directly. This is not as simple for $\mu_1, \ldots, \mu_k$ because the matrix $D(z)$ depends on some vector $z \in M$. More precisely, if we define the matrix D' by $D'_{h,j} = -v_h^T J v_j$, then we have

$$D(z)_{h,j} = D'_{h,j} \cdot (zz^T)_{h,j}$$

for every $h, j \in \{1, \dots, k\}$. Equivalently, we can write $D(z) = D' \circ zz^T$, where the *Hadamard product* of two matrices $A, B \in \mathbb{R}^k$ is defined by

$$A \circ B \in \mathbb{R}^{k \times k}, \ (A \circ B)_{h,j} = A_{h,j} B_{h,j}$$

for $h, j \in \{1, \dots, k\}$. Before we continue investigating the eigenvalues of $D(z)$ we recall the *Rayleigh-Ritz principle*. Usually, the Rayleigh-Ritz principle is stated as a variational formulation of the largest or smallest eigenvalue of a symmetric (or more general Hermitian) matrix (see [54]). We can adapt the corresponding proof to show a similar formulation for the eigenvalue of a skew-symmetric matrix with largest absolute value. Even though this idea is rather straightforward, it seems to be only rarely stated in the literature. Therefore, we provide a detailed proof here.

Lemma 5.2.5. *Let $A \in \mathbb{R}^k$ be a skew-symmetric matrix and let μ_1 be an eigenvalue of A with maximal absolute value. Then:*

$$|\mu_1| = \max_{x^* x = 1} |x^* A x| = \max_{x \in \mathbb{C}^k \setminus \{0\}} \left| \frac{x^* A x}{x^* x} \right|$$

Proof. The second equality is obvious. For the first equality we recall that A is unitarily diagonalizable due to Lemma 5.2.1 (*iii*). Therefore, we can find an orthonormal basis $\{u_1, \dots, u_k\}$ of $\mathbb{C}^k$ such that the basis elements $u_1, \dots, u_k$ are the eigenvectors of A. Let $x \in \mathbb{C}^k$ be a unit vector, i.e. $x^* x = 1$. We can express x in terms of the basis $\{u_1, \dots, u_k\}$. So, there are complex numbers $\alpha_1, \dots, \alpha_k$ such that

$$x = \sum_{j=1}^{k} \alpha_j u_j.$$

Together with the unit length of x this implies:

$$\begin{aligned} 1 = x^* x = \left(\sum_{h=1}^{k} \alpha_h u_h \right)^* \left(\sum_{j=1}^{k} \alpha_j u_j \right) &= \sum_{h=1}^{k} \sum_{j=1}^{k} \bar{\alpha_h} \alpha_j u_h^* u_j \\ &= \sum_{j=1}^{k} \bar{\alpha_j} \alpha_j = \sum_{j=1}^{k} |\alpha_j|^2. \end{aligned}$$

We can use this to bound $|\mu_1|$ from below. To this end, we let $\mu_1, \dots, \mu_k$ be the eigenvalues of A such that u_j is an eigenvector for the eigenvalue μ_j for every $j \in \{1, \dots, k\}$.

$$\begin{aligned} |x^* A x| = \left| \sum_{h,j=1}^{k} \bar{\alpha_h} \alpha_j u_h^* A u_j \right| = \left| \sum_{h,j=1}^{k} \bar{\alpha_h} \alpha_j u_h^* \mu_j u_j \right| &= \left| \sum_{j=1}^{k} |\alpha_j|^2 \mu_j \right| \\ &\leqslant \sum_{j=1}^{k} |\alpha_j|^2 |\mu_j| \leqslant |\mu_1|. \end{aligned}$$

This holds for every unit vector $x \in \mathbb{C}^k$. Therefore, we have

$$|\mu_1| \geqslant \max_{x^* x = 1} |x^* A x|. \tag{5.13}$$

We complete the proof with a simple calculation that implies the equality in (5.13):

$$|\mu_1| = |u_1^* \mu_1 u_1| = |u_1^* A u_1| \leqslant \max_{x^* x = 1} |x^* A x|. \qquad \square$$

We want to study the eigenvalues of the Hadamard product $D' \circ zz^T$. In 1911 Schur [91] investigated such Hadamard products. One of his results is an upper bound on the spectral norm of $A \circ B$, where one of the two matrices A, B is positive semidefinite. We can use this approach to give an upper bound on the largest absolute value of the eigenvalues of $D(z)$. More precisely, because $D(z)$ is skew-symmetric it immediately follows from Lemma 5.2.5 that its spectral norm is

$$||D(z)||_2 := \max_{x^* x = 1} |x^* D(z) x| = \max\{|\mu| \colon \mu \text{ is an eigenvalue of } D(z)\}.$$

Schur formulates his results in the language of bilinear forms. For a translation to the language of matrices, see [58]. While Schur's result considers a Hadamard product $A \circ B$, where A or B is positive semidefinite, we focus on a special case that fits our problem setting. This simplifies the proof.

Theorem 5.2.6. *Let $A \in \mathbb{R}^{k \times k}$ be a skew-symmetric matrix and let $x \in \mathbb{R}^k$. Furthermore, we let μ_1' be an eigenvalue of A and μ_1 be an eigenvalue of $A \circ xx^T$ such that*

$$\mu_1' = \max\{|\mu| \colon \mu \text{ is an eigenvalue of } A\},$$
$$\mu_1 = \max\left\{|\mu| \colon \mu \text{ is an eigenvalue of } A \circ xx^T\right\}.$$

Then:

$$|\mu_1| \leqslant |\mu_1'| \max\left\{x_1^2, \ldots, x_k^2\right\}.$$

Proof. First, we note that $A \circ xx^T$ is skew-symmetric because

$$(A \circ xx^T)^T = A^T \circ (xx^T)^T = -A \circ xx^T.$$

So, by Lemma 5.2.5 we have

$$|\mu_1| = \max\{|u^*(A \circ xx^T)u| \colon u \in \mathbb{C}^k,\ u^* u = 1\}.$$

For every unit vector $u \in \mathbb{C}^k$ we have:

$$\begin{aligned}
|u^*(A \circ xx^T)u| &= \left| \sum_{h=1}^{k} \sum_{j=1}^{k} \bar{u_h} A_{h,j} x_h x_j u_j \right| \\
&= |\tilde{x}^* A \tilde{x}|, \quad \text{with } \tilde{x} = u \circ x \\
&= \left| \frac{\tilde{x}^*}{\tilde{x}^* \tilde{x}} A \frac{\tilde{x}}{\tilde{x}^* \tilde{x}} \right| \cdot (\tilde{x}^* \tilde{x})^2 \\
&\leqslant \max_{v^* v = 1} |v^* A v| \cdot \left(\sum_{j=1}^{k} \bar{u}_j u_j x_j^2 \right) \\
&\leqslant |\mu_1'| \cdot \left(\sum_{j=1}^{k} \bar{u}_j u_j \cdot \max\{x_1^2, \ldots, x_k^2\} \right) \\
&= |\mu_1'| \cdot \underbrace{u^* u}_{=1} \cdot \max\{x_1^2, \ldots, x_k^2\}.
\end{aligned}$$

We note that we require $\widetilde{x} \neq 0$ in the third line of this equation. If this is not the case then $\widetilde{x} = 0$ implies

$$|u^*(A \circ xx^T)u| = |\widetilde{x}^* A \widetilde{x}| = 0 \leqslant |\mu_1'| \max\{x_1^2, \ldots, x_k^2\}. \qquad \square$$

Theorem 5.2.6 bounds an eigenvalue of a matrix $A \circ xx^T$ by using the diagonal elements of xx^T. This is reminiscent of another theorem by Schur [90]: Let $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$. Then

$$\left\{\operatorname{diag} A \colon A \in \mathbb{C}^{k\times k} \text{ Hermitian with eigenvalues } \lambda_1, \ldots, \lambda_k\right\}$$

is contained in the convex hull of

$$\left\{ \left(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(k)}\right)^T : \sigma \in \operatorname{Sym}_k \right\}.$$

Conversely, Horn [53] proved that for every point x in this convex hull there is a Hermitian matrix A with eigenvalues $\lambda_1, \ldots, \lambda_k$ such that $x = \operatorname{diag} A$. As it turns out, these results have a generalization that can be related to symplectic geometry [13].

We now have all the pieces together, to derive an upper bound on $1/4c_{\mathrm{EHZ}}(C)$. For this, we keep in mind that every eigenvalue of a skew-symmetric matrix is purely imaginary. So, its absolute value equals its imaginary part if the imaginary part is nonnegative.

$$\begin{aligned}
\frac{1}{4c_{\mathrm{EHZ}}(C)} &\overset{(5.5)}{=} \frac{1}{2} \max_{z\in M} \left(\max_{X\in\Pi_k} \left\langle F, X^T D(z) X \right\rangle \right) \\
&\overset{(5.7)}{=} \frac{1}{2} \max_{z\in M} \left(\max_{X\in\Pi_k} \left\langle F', X^T D(z) X \right\rangle \right) \\
&\overset{(5.11)}{\leqslant} \frac{1}{2} \max_{z\in M} \left(- \min_{\sigma\in\operatorname{Sym}_k} \sum_{j=1}^{k} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_{\sigma(j)} \right) \qquad (5.14) \\
&\overset{(5.12)}{=} \max_{z\in M} \left(\sum_{j=1}^{\lfloor \frac{k}{2} \rfloor} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_j \right) \\
&\leqslant \max_{z\in M} \left(\sum_{j=1}^{\lfloor \frac{k}{2} \rfloor} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_1' \cdot \max\left\{z_1^2, \ldots, z_k^2\right\} \right),
\end{aligned}$$

where μ_1' is an eigenvalue of D' with maximum absolute value. The last line in this equation follows from Theorem 5.2.6 and the fact that $\operatorname{Im}\mu_1 \geqslant \operatorname{Im}\mu_j$ for every $j \in \{1, \ldots, k\}$. To compute the upper bound given by (5.14) we note that $\mu_1', \lambda_1, \ldots, \lambda_k$ do not depend on the choice of z. Thus, it remains to solve the following problem:

$$\begin{aligned}
\max\ & z_h^2 \\
\text{s. t. } & \sum_{j=1}^{k} z_j v_j = 0 \\
& \sum_{j=1}^{k} z_j = 1 \qquad (5.15) \\
& z_j \geqslant 0 \ \ \forall\, j \in [k] \\
& h \in \{1, \ldots, k\}.
\end{aligned}$$

We can solve this problem with k LPs. For this, we remove the last constraint, successively solve the resulting LP for every $h \in \{1, \dots, k\}$ and take the largest optimal value. We remark that these problems are indeed LPs even though their objective functions are quadratic. Since the variables $z_1, \dots, z_k$ are nonnegative, we can maximize z_h first and square afterwards.

In Table 6.3 we give some numerical results for this upper bound. Like in Chapter 4 the input is a random polytope C that we get by taking the convex hull of normally distributed random points. The polar set C° is part of the input as well. Additionally, we provide numerical results for larger instances in Table 6.4. We can see from these tables that the approach in this chapter provides a (computationally) cheap upper bound but the quality of this bound is rather poor, especially for larger instances, i.e. where the number of facets k is big.

One could suspect that the reason for the poor quality of this upper bound lies within the last inequality in (5.14). There, we make the estimate

$$\operatorname{Im}\mu_j \leqslant \operatorname{Im}\mu_1 \quad \text{for all } j \in \{1, \dots, k\},$$

where $\mu_1, \dots, \mu_k$ are the eigenvalues of $D(z)$, ordered by their imaginary part. It turns out that this estimate is indeed one reason, but not necessarily the only explanation for the gap between our upper bound and the exact value of $1/4c_{\mathrm{EHZ}}(C)$. More precisely, for a given polytope C we can pick a large number of points in

$$M = \left\{ z \in \mathbb{R}^k_{\geqslant 0} \colon \sum_{j=1}^{k} z_j v_j = 0,\ \sum_{j=1}^{k} z_j = 1 \right\}$$

and compute the corresponding eigenvalues of $D(z)$ for each choice. With this approach we can approximate

$$\max_{z \in M} \left(\sum_{j=1}^{\lfloor \frac{k}{2} \rfloor} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_j \right),$$

so we avoid the last inequality in (5.14). As an example, we consider line 17 in Table 6.3 in the appendix. There, we examine a randomly generated polytope C for which we find that the exact value of $1/4c_{\mathrm{EHZ}}(C)$ (rounded to eight decimal places) is 0.01286290 and our upper bound is 1.702583848. If we avoid the last inequality in (5.14) we get the improved upper bound

$$\max_{z \in M} \left(\sum_{j=1}^{\lfloor \frac{k}{2} \rfloor} \operatorname{Im}\lambda_j \cdot \operatorname{Im}\mu_j \right) \approx 0.090131495\,.$$

Another example, that we give here, is the square

$$C = \operatorname{conv}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}.$$

The exact value of $1/4c_{\mathrm{EHZ}}(C)$ is 0.0625 and our upper bound is 0.70710678. The improved upper bound is 0.15088835.

5.3 Upper bounds via semidefinite relaxation

Another approach to compute upper bounds on $1/4c_{\mathrm{EHZ}}(C)$ is to formulate (5.4) as a CP (see (3.7) in Chapter 3) and to employ the strategy that we describe in Chapter 3.2.4. More precisely, we relax the complete positivity constraint to get an SDP that we can solve. As in the previous chapter we have to discuss how to find an optimal permutation σ. One way is to relax the fact that $v_1, \ldots, v_k$ are vertices of C°. This results in a single CP that we investigate. Another way is to fix the permutation σ in (5.4) and solve $k!$ maximization problems. The latter approach is clearly only applicable for small values of k. On the upside we only relax (5.4) once, namely when we go from completely positive to semidefinite optimization. Thus, we can expect sharper bounds with the latter approach.

5.3.1 A completely positive formulation of the EHZ capacity

Consider the following quadratic optimization problem:

$$\begin{aligned} \max \quad & -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} y_i y_j w_i^T J w_j \\ \text{s. t.} \quad & \sum_{i=1}^{k} y_i w_i = 0 \\ & \sum_{i=1}^{k} y_i = 1 \\ & y_i \geqslant 0 \ \ \forall\, i \in \{1, \ldots, k\} \\ & w_1, \ldots, w_k \in C^\circ. \end{aligned} \tag{5.16}$$

If we let $(y_1, \ldots, y_k, \sigma)$ be an optimal solution of (5.4), then $(y_1, \ldots, y_k, v_{\sigma(1)}, \ldots, v_{\sigma(k)})$ is a feasible solution for (5.16), where $v_1, \ldots, v_k$ are the vertices of C°. Thus, the optimal value of (5.16) is an upper bound on $1/4c_{\mathrm{EHZ}}(C)$.

To simplify (5.16) we recall a property of the polar polytope C°. We let W be a matrix whose rows are exactly the vertices of C. Then, since $0 \in \operatorname{int} C$, we have

$$C^\circ = \{x\colon Wx \leqslant e\}, \tag{5.17}$$

where $e = (1, \ldots, 1)^T$. For a proof, see [105]. We use (5.17) to prove a more general property.

Lemma 5.3.1. *Let $C \subseteq \mathbb{R}^n$ be a polytope with $0 \in \operatorname{int} C$ and let W be a matrix whose rows are exactly the vertices of C. Then for every $\alpha \in \mathbb{R}$:*

$$\alpha C^\circ = \{x\colon Wx \leqslant \alpha e\}.$$

Proof. For $\alpha \neq 0$ the claim immediately follows from (5.17):

$$\alpha C^\circ = \{\alpha x\colon Wx \leqslant e\} = \left\{x\colon \frac{1}{\alpha} Wx \leqslant e\right\} = \{x\colon Wx \leqslant \alpha e\}.$$

Now we let $\alpha = 0$. It is obvious that

$$0 \cdot C^\circ = \{0\} \subseteq \{x \colon Wx \leqslant 0\}.$$

Thus, it remains to prove that $Wx \leqslant 0$ implies $x = 0$. Assume this is false, i.e. there is $x_0 \in \mathbb{R}^n$ with $x_0 \neq 0$ and $Wx_0 \leqslant 0$. This means that $v^T x_0 \leqslant 0$ holds for every vertex v of C. Therefore, the halfspace

$$H^- = \{z \in \mathbb{R}^n \colon z^T x_0 \leqslant 0\}$$

contains C and hence $\operatorname{int} C \subseteq \operatorname{int} H^-$. This contradicts the fact that $0 \in \operatorname{int} C$ but $0 \notin \operatorname{int} H^-$. It follows that $x = 0$ if and only if $Wx \leqslant 0$ which concludes the proof. □

In the objective function and in the first contraint of (5.16) the quadratic terms of the form $y_i w_i$ appear, where $y_i \geqslant 0$ and $w_i \in C^\circ$. We substitute $z_i := y_i w_i$ and require $z_i \in y_i C^\circ$. With Lemma 5.3.1 we get the following quadratic maximization problem:

$$\begin{aligned} \max \ & -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{i-1} z_i^T J z_j \\ \text{s. t. } & \sum_{i=1}^{k} z_i = 0 \\ & \sum_{i=1}^{k} y_i = 1 \\ & Wz_i \leqslant y_i e \quad \forall i \in \{1, \ldots, k\} \\ & y_i \geqslant 0 \quad \forall i \in \{1, \ldots, k\}. \end{aligned} \tag{5.18}$$

According to Burer [25] we can reformulate every quadratic optimization problem with binary and/or continuous variables as a CP. The problem that we are concerned with, i.e. (5.18), has a convenient form which we can exploit to find such a reformulation in a particularly simple way. In the following we deduce the corresponding CP for this special case of quadratic programs and give a proof that is independent of [25]. First, we need some basic convex geometry statements. To this end, we recall the definitions of the completely positive and the copositive cone from Chapter 3.2.4:

$$\begin{aligned} \mathcal{CP}_n &:= \operatorname{cone} \left\{ xx^T \colon x \in \mathbb{R}^n_{\geqslant 0} \right\}, \\ \mathcal{COP}_n &:= \left\{ A \in \mathcal{S}^n \colon x^T A x \geqslant 0 \ \ \forall\, x \in \mathbb{R}^n_{\geqslant 0} \right\}. \end{aligned}$$

Lemma 5.3.2. *Let $A \in \mathcal{COP}_n$ be a copositive matrix. Then:*

$$\{X \in \mathcal{CP}_n \colon \langle A, X \rangle = 0\} = \operatorname{cone}\{xx^T \colon x^T A x = 0,\ x \geqslant 0\}.$$

Proof. It is obvious that the inclusion "$\supseteq$" holds since $x^T A x = \left\langle A, xx^T \right\rangle$. To show the converse, take $X \in \mathcal{CP}_n$ with $\langle A, X \rangle = 0$. By definition we can write

$$X = \sum_{i=1}^{N} \alpha_i x_i x_i^T,$$

where $x_i \in \mathbb{R}^n_{\geqslant 0}$, $\alpha_i \in \mathbb{R}_{\geqslant 0}$ and $N \in \mathbb{N}$. We assume that $\alpha_1, \ldots, \alpha_N$ are positive. Otherwise, we drop every term where $\alpha_i = 0$. We get:

$$0 = \left\langle A, \sum_{i=1}^{N} \alpha_i x_i x_i^T \right\rangle = \sum_{i=1}^{N} \alpha_i x_i^T A x_i.$$

Since A is copositive, every term of the last sum is nonnegative. Because the sum is 0, it follows that every term has to be 0. Furthermore, $\alpha_1, \ldots, \alpha_N > 0$ implies $x_i^T A x_i = 0$ for every $i \in \{1, \ldots, N\}$. Thus, the inclusion "$\subseteq$" holds as well. □

Lemma 5.3.3. *[16] Let $C \subseteq \mathbb{R}^n$ be a convex, closed set that does not contain straight lines. Then for every $c \in \mathbb{R}^n$*

$$\max\{c^T x \colon x \in C\}$$

is either ∞ or attained at an extreme point of C.

We omit the proof of Lemma 5.3.3 since it is a well-established fact. It immediately follows from Theorem 3.2 and Lemma 3.5 in [16].

Lemma 5.3.4. *Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a cone and let $H \subseteq \mathbb{R}^n$ be a hyperplane. Then every extreme point of $\mathcal{K} \cap H$ lies in an extreme ray of $\mathcal{K}$.*

Proof. We prove the statement via contradiction. Let x be an extreme point of $\mathcal{K} \cap H$. Assume there are $v, w \in \mathcal{K}$ such that $x = (v + w)/2$ and that v, w are not multiples of x. Define the set

$$V_x = \{u \in \mathbb{R}^n \colon \exists\, \varepsilon > 0 \text{ with } x + \varepsilon u, x - \varepsilon u \in \mathcal{K}\}.$$

It is easy to check that V_x is a linear subspace of $\mathbb{R}^n$ and that $x, v - x \in V_x$. Note that x and v are linearly independent and thus $\dim V_x \geqslant 2$. We let H_0 be the linear space that we get by shifting the hyperplane H. In particular, it is $\dim H_0 = n - 1$. Consequently, we have

$$\dim(V_x \cap H_0) \geqslant \dim V_x - 1 \geqslant 1.$$

So, there is a nonzero vector z that lies in both V_x and H_0. On the one hand, this means there is $\varepsilon > 0$ such that

$$\begin{aligned} z_1 &:= x + \varepsilon z \in \mathcal{K}, \\ z_2 &:= x - \varepsilon z \in \mathcal{K}, \\ x &= \frac{1}{2}(z_1 + z_2). \end{aligned}$$

On the other hand, if we let $H = \{y \colon a^T y = b\}$ and $H_0 = \{y \colon a^T y = 0\}$, then

$$\begin{aligned} a^T z_1 &= a^T x + \varepsilon a^T z = b, \\ a^T z_2 &= a^T x - \varepsilon a^T z = b. \end{aligned}$$

Thus, $z_1, z_2 \in \mathcal{K} \cap H$. The fact that x is an extreme point of $\mathcal{K} \cap H$ implies $z_1 = z_2$ and therefore $z = 0$. This contradicts the choice of $z \neq 0$ and the claim follows. □

Putting the statements in the previous three lemmas together, we derive a way to formulate every quadratic optimization problem of a certain type as a CP.

Theorem 5.3.5. *Let $A \in \mathbb{R}^{n\times m}$, $Q \in \mathbb{R}^{m\times m}$ and $c \in \mathbb{R}^m_{\geqslant 0}$. Then:*

$$\begin{array}{lll} \max & x^TQx & = \\ \text{s. t.} & Ax = 0 & \\ & c^Tx = 1 & \\ & x \in \mathbb{R}^m_{\geqslant 0} & \end{array} \qquad \begin{array}{lll} \max & \langle Q, X\rangle \\ \text{s. t.} & \left\langle A^TA, X\right\rangle = 0 \\ & \left\langle cc^T, X\right\rangle = 1 \\ & X \in \mathcal{CP}_m, \end{array}$$

if both maxima are finite.

Proof. First, we rewrite the quadratic problem on the left-hand side such that its constraints are quadratic as well. For the first constraint we note that $Ax = 0$ is equivalent to $(Ax)^T(Ax) = 0$. Since we require that x and c are nonnegative, the second constraint is equivalent to

$$1 = (x^Tc)^2 = (x^Tc)(x^Tc)^T = x^Tcc^Tx.$$

We can write the nonnegativity constraint as $x_ix_j \geqslant 0$ for every $i, j \in \{1, \dots, m\}$. This implies that $x \geqslant 0$ or $x \leqslant 0$. The problem that we get contains only quadratic terms and no linear terms. Thus, for every solution $x \leqslant 0$ we have that $-x$ is a solution as well and both x and $-x$ have the same objective value. Therefore, we get:

$$\begin{array}{lll} \max & x^TQx & = \\ \text{s. t.} & Ax = 0 & \\ & c^Tx = 1 & \\ & x \in \mathbb{R}^m_{\geqslant 0} & \end{array} \qquad \begin{array}{lll} \max & x^TQx \\ \text{s. t.} & x^TA^TAx = 0 \\ & x^Tcc^Tx = 1 \\ & x_ix_j \geqslant 0 \ \ \forall\, i, j \in \{1, \dots, k\}. \end{array} \tag{5.19}$$

We prove that the maximum on the right-hand side of (5.19) is equal to the optimal value of the CP in the claim of the theorem. It is easy to see that the optimal value of the CP is at least as big as the maximum in (5.19). For every feasible solution $x \in \mathbb{R}^m_{\geqslant 0}$ we can take $X = xx^T$ and observe that X is a feasible solution for the CP with objective value x^TQx.

Now we show the converse. Application of Lemma 5.3.2 yields that the set of feasible solutions of the CP is the intersection of a cone $\mathcal{K}$ with a hyperplane H. More precisely,

$$\begin{aligned} \mathcal{K} &= \text{cone}\{xx^T : x^TA^TAx = 0,\ x \in \mathbb{R}^m_{\geqslant 0}\}, \\ H &= \{X \in \mathcal{S}^m : \left\langle cc^T, X\right\rangle = 1\}. \end{aligned}$$

Note that the matrix A^TA is positive semidefinite and hence copositive. $\mathcal{K} \cap H$ is convex and closed because both $\mathcal{K}$ and H are convex and closed as well. Moreover, $K \cap H$ does not contain straight lines because it is contained in the proper cone

$$\{X \in \mathcal{S}^m : X_{i,j} \geqslant 0 \ \ \forall\, i, j \in \{1, \dots, m\}\}.$$

If we identify $\mathcal{S}^m$ with $\mathbb{R}^{m(m+1)/2}$, we can apply Lemma 5.3.3. Thus, if the optimal value of the CP is finite, it is attained at an extreme point X^* of $\mathcal{K} \cap H$. Due to Lemma 5.3.4 this

extreme point X^* lies in an extreme ray of $\mathcal{K}$. So, X^* has the form xx^T with some $x \in \mathbb{R}^m_{\geqslant 0}$ such that $x^T A^T A x = 0$. It is straightforward to check that x is feasible for the quadratic problem on the right-hand side of (5.19) and that its objective value equals $\langle Q, X^* \rangle$. □

We can use Theorem 5.3.5 to formulate (5.18) as a CP. To this end, we bring (5.18) into a form that fits the requirements of Theorem 5.3.5. We introduce slack variables $s_1, \ldots, s_k \in \mathbb{R}^k_{\geqslant 0}$ and for every $i \in \{1, \ldots, k\}$ we write $z_i = z_i^+ - z_i^-$ with $z_i^+, z_i^- \in \mathbb{R}^{2n}_{\geqslant 0}$. Then we get:

$$\begin{aligned}
(5.18) \;=\; \max \;\; & -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{i-1} \left(z_i^{+T} J z_j^+ - z_i^{+T} J z_j^- - z_i^{-T} J z_j^+ + z_i^{-T} J z_j^- \right) \\
\text{s. t.} \;\; & \sum_{i=1}^{k} \left(z_i^+ - z_i^- \right) = 0 \\
& \sum_{i=1}^{k} y_i = 1 \\
& W z_i^+ - W z_i^- - y_i e + s_i = 0 \quad \forall\, i \in \{1, \ldots, k\} \\
& y_i, z_i^+, z_i^-, s_i \geqslant 0 \quad \forall\, i \in \{1, \ldots, k\}.
\end{aligned} \tag{5.20}$$

Now we can apply Theorem 5.3.5 and obtain a CP. Afterwards, we perform a semidefinite relaxation as described in Chapter 3.2.4. More precisely, we replace the constraint $X \in \mathcal{CP}_m$ by $X \in \mathcal{S}^m_{\geq 0} \cap \mathbb{R}^{m \times m}_{\geqslant 0}$. This results in the following SDP whose optimal value is an upper bound on $1/4c_{\mathrm{EHZ}}(C)$:

$$\begin{aligned}
\frac{1}{4c_{\mathrm{EHZ}}(C)} \;\leqslant\; \max \;\; & \langle Q, X \rangle \\
\text{s. t.} \;\; & \left\langle A^T A, X \right\rangle = 0 \\
& \left\langle cc^T, X \right\rangle = 1 \\
& X_{i,j} \geqslant 0, \; \forall\, i, j \in \{1, \ldots, \ell\} \\
& X \in \mathcal{S}^\ell_{\geq 0},
\end{aligned} \tag{5.21}$$

where m is the number of vertices of C, k is the number of facets of C and $\ell = k(1+4n+k)$. Here, we let the matrices $Q \in \mathbb{R}^{\ell \times \ell}, A \in \mathbb{R}^{2n+km \times \ell}$ and the vector $c \in \mathbb{R}^\ell_{\geqslant 0}$ be such that the quadratic optimization problem (5.20) has the same form as in Theorem 5.3.5.

In Table 6.5 we provide numerical results for this upper bound. We observe that the upper bound is much better than the one that we derive in Chapter 5.2 and that the required running time is still reasonable. However, we run out of memory rather quickly because the input for the SDP in (5.21), i.e. Q, A and c, is rather large, even for polytopes with dimension $\leqslant 4$ and with few vertices.

5.3.2 A more accurate bound with multiple semidefinite programs

As mentioned earlier, another way to handle the maximization in (5.4) is to remove the constraint $\sigma \in \mathrm{Sym}_k$ and to solve the resulting problem for every permutation σ. In comparison to the approach in Chapter 5.3.1, this method prevents that we run out of memory

quickly at the cost of longer running time. Thus, for $\sigma \in \mathrm{Sym}_k$ we focus on the following problem:

$$\begin{aligned} p_\sigma^* := \max \quad & -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} y_i y_j v_{\sigma(i)}^T J v_{\sigma(j)} \\ \text{s. t.} \quad & \sum_{i=1}^{k} y_i v_{\sigma(i)} = 0 \\ & \sum_{i=1}^{k} y_i = 1 \\ & y_i \geqslant 0 \ \ \forall\, i \in \{1,\dots,k\}, \end{aligned} \tag{5.22}$$

where $v_1,\dots,v_k$ are the vertices of C°. From (5.4) we have

$$\max\{p_\sigma^* \colon \sigma \in \mathrm{Sym}_k\} = \frac{1}{4c_{\mathrm{EHZ}}(C)}.$$

We notice that the maximization problem (5.22) is a quadratic optimization problem that has the same form as the quadratic problem in Theorem 5.3.5. Moreover, p_σ^* is finite because the set of feasible solutions in (5.22) is contained in the standard simplex

$$\Delta_{k-1} := \{x \in \mathbb{R}^k_{\geqslant 0} \colon e^T x = 1\},$$

which is bounded. So, we can apply Theorem 5.3.5 and get a completely positive formulation of p_σ^*:

$$\begin{aligned} p_\sigma^* := \max \quad & \langle Q, X\rangle \\ \text{s. t.} \quad & \left\langle A^T A, X\right\rangle = 0 \\ & \left\langle ee^T, X\right\rangle = 1 \\ & X \in \mathcal{CP}_k, \end{aligned} \tag{5.23}$$

where A is the $(2n \times k)$-matrix whose ith column is $v_{\sigma(i)}$ for every $i \in \{1,\dots,k\}$ and the matrix $Q \in \mathbb{R}^{k\times k}$ is given by

$$Q_{i,j} = \begin{cases} -\frac{1}{2} v_{\sigma(i)} J v_{\sigma(j)}, & \text{if } i > j, \\ 0, & \text{if } i \leqslant j, \end{cases}$$

for every $i, j \in \{1,\dots,k\}$. We remark that strong duality holds for (5.23), i.e. it is $p_\sigma^* = d_\sigma^*$ for every $\sigma \in \mathrm{Sym}_k$, where

$$\begin{aligned} d_\sigma^* := \min \quad & s \\ \text{s. t.} \quad & r, s \in \mathbb{R} \\ & rA^T A + see^T - Q \in \mathcal{COP}_k \end{aligned} \tag{5.24}$$

is the optimal value of the dual program of (5.23). This strong duality follows from Theorem 3.2.3 (iv) if (5.24) is bounded and has a strictly feasible solution. We know from (5.17) that the polar polytope C° contains 0. So, there is $\alpha \in \mathbb{R}^k_{\geqslant 0}$ such that

$$0 = \sum_{i=1}^{k} \alpha_i v_{\sigma(i)} \quad \text{and} \quad 1 = \sum_{i=1}^{k} \alpha_i.$$

This means that $\alpha\alpha^T$ is a feasible solution of (5.23). With Theorem 3.2.3 (i) we find that $d^*_\sigma > -\infty$. On the other hand, we get a strictly feasible solution for (5.24) if we pick

$$r \geqslant 0 \quad \text{and} \quad s > \max\{Q_{i,j} \colon i, j \in \{1, \ldots, k\}\}.$$

The fact that $A^T A$ is positive semidefinite implies

$$x^T(rA^TA + see^T - Q)x = rx^TA^TAx + x^T(see^T - Q)x \geqslant \sum_{i=1}^{k} \sum_{j=1}^{i-1} x_i x_j (s - Q_{i,j}) > 0,$$

for every nonzero $x \in \mathbb{R}^k_{\geqslant 0}$. With Theorem 3.2.10 we see that (r, s) is indeed a strictly feasible solution of (5.24) and strong duality follows.

We handle the CP (5.23) in the same way as in Chapter 5.3.1. More precisely, we relax the complete positivity constraint to get

$$\begin{aligned} \widehat{p}_\sigma := \max \ & \langle Q, X \rangle \\ \text{s. t.} \ & \langle A^TA, X \rangle = 0 \\ & \langle ee^T, X \rangle = 1 \\ & X_{i,j} \geqslant 0 \ \ \forall\, i, j \in \{1, \ldots, k\} \\ & X \in \mathcal{S}^k_{\geq 0}. \end{aligned} \tag{5.25}$$

Thus, we obtain an upper bound on $1/4c_{\mathrm{EHZ}}(C)$ due to

$$\frac{1}{4c_{\mathrm{EHZ}}(C)} = \max\{p^*_\sigma \colon \sigma \in \mathrm{Sym}_k\} \leqslant \max\{\widehat{p}_\sigma \colon \sigma \in \mathrm{Sym}_k\}. \tag{5.26}$$

To compute this bound we need to solve $k!$ SDPs. However, we can reduce this number from $k!$ to $(k-1)!$. To this end, we note that if $x \colon [0,1] \to \mathbb{R}^{2n}$ is a feasible solution of the optimization problem in Theorem 2.3.6, then $x \circ \varphi$ is also a feasible solution and $\mathbb{A}(x) = \mathbb{A}(x \circ \varphi)$, where

$$\varphi \colon [0,1] \to [0,1], \ \ \varphi(t) = \begin{cases} t + a, & t \in [0, 1-a], \\ t + a - 1, & t \in (1-a, 1], \end{cases}$$

for some $a \in [0,1]$. Note that $x \circ \varphi$ is a continuous curve because of $x(0) = x(1)$ and that this curve is closed. More precisely,

$$x \circ \varphi(0) = 0 + a = 1 + a - 1 = x \circ \varphi(1).$$

Showing that $x \circ \varphi$ is indeed a feasible solution of the maximization problem in Theorem 2.3.6 is straightforward. Additionally, we have

$$\begin{aligned}\mathbb{A}(x \circ \varphi) &= -\frac{1}{2}\left[\int_0^{1-a} \nabla(x \circ \varphi(t))^T J(x \circ \varphi)(t) \cdot \varphi'(t)\, dt \right.\\ &\qquad \left. + \int_{1-a}^{1} \nabla(x \circ \varphi)(t)^T J(x \circ \varphi)(t) \cdot \varphi'(t)\, dt\right]\\ &= -\frac{1}{2}\left[\int_a^1 \nabla x(t)^T J x(t)\, dt + \int_0^a \nabla x(t)^T J x(t)\, dt\right]\\ &= -\frac{1}{2}\int_0^1 \nabla x(t)^T J x(t)\, dt = \mathbb{A}(x).\end{aligned}$$

The first line of this equation holds since $\varphi'(t) = 1$ for every $t \in (0, 1-a)$ and every $t \in (1-a, 1)$. Thus, if we let $\sigma \in \mathrm{Sym}_k$ be such that $p^*_\sigma = 1/4c_{\mathrm{EHZ}}(C)$, then we also have

$$p^*_{\sigma\circ\tau} = \frac{1}{4c_{\mathrm{EHZ}}(C)},$$

for every circular shift τ. Here, we say that a permutation τ is a *circular shift* if there is $a \in \mathbb{Z}$ such that

$$\tau(i) = (i + a) \bmod k$$

for every $i \in \{1, \ldots, k\}$. We want to prove that the same principle also holds for our upper bounds, i.e. that $\widehat{p}_\sigma = \widehat{p}_{\sigma\circ\tau}$ holds for every $\sigma, \tau \in \mathrm{Sym}_k$, where τ is a circular shift. Before we prove this result we show the following lemma.

Lemma 5.3.6. *Let $A \in \mathbb{R}^{n\times n}$ be a positive semidefinite matrix with*

$$\sum_{i=1}^{k}\sum_{j=1}^{k} A_{i,j} = 0.$$

Then every row and column sum of A is equal to zero.

Proof. Due to the symmetry of A it suffices to show that the row sums are equal to zero. Since A is positive semidefinite, it has a Cholesky decomposition $A = LL^T$ and we get

$$0 = e^T A e = e^T L L^T e = ||L^T e||^2.$$

This implies that $L^T e = 0$ and therefore

$$Ae = LL^T e = 0.$$

□

Theorem 5.3.7. *Let $C \subseteq \mathbb{R}^{2n}$ be a polytope with $0 \in \operatorname{int} C$ and let $v_1, \ldots, v_k$ be the vertices of C°. Furthermore, let $\sigma, \tau \in \mathrm{Sym}_k$ and let τ be a circular shift. Then $\widehat{p}_\sigma = \widehat{p}_{\sigma\circ\tau}$.*

Proof. As before we let $A \in \mathbb{R}^{2n\times k}$ be the matrix whose ith column is $v_{\sigma(i)}$ and we let $Q \in \mathbb{R}^{k\times k}$ be defined by

$$Q_{i,j} = \begin{cases} -\frac{1}{2} v_{\sigma(i)} J v_{\sigma(j)}, & \text{if } i > j, \\ 0, & \text{if } i \leqslant j. \end{cases}$$

To simplify the notation in this proof we let $Q' = 2(Q^T - Q)$. We note that

$$Q'_{i,j} = v_{\sigma(i)}^T J v_{\sigma(j)}$$

holds for every $i, j \in \{1, \ldots, k\}$. Furthermore, we define $w_1, \ldots, w_{2k} \in \mathbb{R}^{2n}$ by

$$w_i = v_{\sigma(i)}\,,\; w_{k+i} = J v_{\sigma(i)}\,,$$

for every $i \in \{1, \ldots, k\}$ and we let $B \in \mathbb{R}^{2k\times 2k}$ be the corresponding Gramian matrix:

$$B = \left(w_i^T w_j\right)_{i,j\in\{1,\ldots,k\}}.$$

B is positive semidefinite and it has the following form:

$$B = \begin{pmatrix} A^T A & Q' \\ Q'^T & A^T A \end{pmatrix}.$$

Next, we let X be an optimal solution for the SDP (5.25). In particular, $\widehat{p}_\sigma = \langle Q, X\rangle$. Since X is positive semidefinite, it has a Cholesky decomposition $X = LL^T$. We define another matrix $D \in \mathbb{R}^{2k}$ by

$$D := \begin{pmatrix} X & X \\ X & X \end{pmatrix} = \begin{pmatrix} LL^T & LL^T \\ LL^T & LL^T \end{pmatrix} = \begin{pmatrix} L \\ L \end{pmatrix} \begin{pmatrix} L \\ L \end{pmatrix}^T \geq 0.$$

The Hadamard product $B \circ D$ is positive semidefinite since both B and D are positive semidefinite [103]. We get

$$0 \leqslant e^T (B \circ D) e = 2 \sum_{i=1}^{k} \sum_{j=1}^{k} \left(X_{i,j} A^T A_{i,j} + X_{i,j} Q'_{i,j} \right) = 2 \sum_{i=1}^{k} \sum_{j=1}^{k} X_{i,j} Q'_{i,j}. \tag{5.27}$$

In the last equation we use $\left\langle X, A^T A \right\rangle = 0$. If we change the signs for $w_{k+1}, \ldots, w_{2k}$, we can apply the exact same reasoning. Note that instead of the matrix B we have the matrix

$$\begin{pmatrix} A^T A & -Q' \\ -Q'^T & A^T A \end{pmatrix}$$

and instead of (5.27) we get:

$$0 \leqslant -2 \sum_{i=1}^{k} \sum_{j=1}^{k} X_{i,j} Q'_{i,j}. \tag{5.28}$$

Together, (5.27) and (5.28) imply $e^T(B \circ D)e = 0$, so all the entries of $B \circ D$ sum up to zero. Due to Lemma 5.3.6 every row sum of $B \circ D$ is equal to zero. For every $i \in \{1, \dots, k\}$ this means

$$\sum_{j=1}^{k} X_{i,j}(A^T A)_{i,j} + \sum_{j=1}^{k} X_{i,j} Q'_{i,j} = 0.$$

Note that $X \circ A^T A$ is a positive semidefinite matrix whose entries sum up to zero because of $\langle A^T A, X \rangle = 0$. So, Lemma 5.3.6 also implies:

$$\sum_{j=1}^{k} X_{i,j}(A^T A)_{i,j} = 0$$

for every $i \in \{1, \dots, k\}$. It follows that

$$\sum_{j=1}^{k} X_{i,j} Q'_{i,j} = 0. \tag{5.29}$$

Now, we construct a feasible solution for the SDP (5.25) that corresponds to $\sigma \circ \tau$. First, we assume that

$$\tau(i) = (i+1) \bmod k$$

for every $i \in \{1, \dots, k\}$. We let $X^\tau \in \mathbb{R}^{k \times k}$ be given by

$$X^\tau_{i,j} = X_{\tau(i),\tau(j)}$$

for every $i, j \in \{1, \dots, k\}$. The goal is now to show that X^τ fulfils the constraints of the SDP for $\sigma \circ \tau$. To this end, we denote the $(2n \times k)$-matrix whose ith column is $v_{\sigma\circ\tau(i)}$ by A^τ and we let $Q^\tau \in \mathbb{R}^{k \times k}$ by given by

$$Q^\tau_{i,j} = \begin{cases} -\frac{1}{2} v_{\sigma\circ\tau(i)} J v_{\sigma\circ\tau(j)}, & \text{if } i > j, \\ 0, & \text{if } i \leqslant j. \end{cases}$$

It is clear that $X^\tau_{i,j} \geqslant 0$ and that $X^\tau \succeq 0$ because we have $X^\tau = P^T X P$ for some permutation matrix P. It is also easy to see that $\langle ee^T, X^\tau \rangle = 1$ because

$$1 = \sum_{i=1}^{k} \sum_{j=1}^{k} X_{i,j} = \sum_{i=1}^{k} \sum_{j=1}^{k} X_{\tau(i),\tau(j)}.$$

Similarly, the remaining constraint $\langle (A^\tau)^T (A^\tau), X^\tau \rangle = 0$ follows from

$$0 = \sum_{i=1}^{k} \sum_{j=1}^{k} v_i^T v_j X_{i,j} = \sum_{i=1}^{k} \sum_{j=1}^{k} v_{\tau(i)}^T v_{\tau(j)} X_{\tau(i),\tau(j)}.$$

Showing that the objective values are equal takes a little bit more effort.

$$\begin{aligned}
\langle Q^\tau, X^\tau \rangle &= -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{i-1} X_{\tau(i),\tau(j)} v_{\sigma\circ\tau(i)}^T J v_{\sigma\circ\tau(j)} \\
&= -\frac{1}{2} \left(\sum_{i=1}^{k-1} \sum_{j=1}^{i-1} X_{\tau(i),\tau(j)} v_{\sigma\circ\tau(i)}^T J v_{\sigma\circ\tau(j)} \right) - \frac{1}{2} \left(\sum_{j=1}^{k-1} X_{\tau(k),\tau(j)} v_{\sigma\circ\tau(k)}^T J v_{\sigma\circ\tau(j)} \right) \\
&= -\frac{1}{2} \left(\sum_{i=2}^{k} \sum_{j=2}^{i-1} X_{i,j} v_{\sigma(i)}^T J v_{\sigma(j)} \right) - \frac{1}{2} \left(\sum_{j=2}^{k} X_{1,j} v_{\sigma(1)}^T J v_{\sigma(j)} \right) \\
&= -\frac{1}{2} \left(\sum_{i=1}^{k} \sum_{j=1}^{i-1} X_{i,j} v_{\sigma(i)}^T J v_{\sigma(j)} \right) - \left(\sum_{j=1}^{k} X_{1,j} v_{\sigma(1)}^T J v_{\sigma(j)} \right) \\
&= \langle Q, X \rangle - \sum_{j=1}^{k} X_{1,j} Q'_{1,j}
\end{aligned}$$

In the second to last line of this equation we add the term

$$-\frac{1}{2} \left(\sum_{i=1}^{k} X_{i,1} v_{\sigma(i)}^T J v_{\sigma(1)} \right) - \frac{1}{2} \left(\sum_{j=1}^{k} X_{1,j} v_{\sigma(1)}^T J v_{\sigma(j)} \right).$$

This term is equal to zero because J is skew-symmetric. Now we can apply (5.29) and obtain that both objective values are equal. This implies $\widehat{p}_\sigma \leqslant \widehat{p}_{\sigma\circ\tau}$. In particular, we find that

$$\widehat{p}_\sigma \leqslant \widehat{p}_{\sigma\circ\tau} \leqslant \widehat{p}_{\sigma\circ\tau^2} \leqslant \ldots \leqslant \widehat{p}_{\sigma\circ\tau^{k-1}} \leqslant \widehat{p}_{\sigma\circ\tau^k} = \widehat{p}_\sigma.$$

Thus, all these inequalities hold with equality. This completes the proof because every circular shift has the form τ^i for some $i \in \{1, \ldots, k\}$. □

Due to Theorem 5.3.7 we do not have to calculate both $\widehat{p}_\sigma$ and $\widehat{p}_\rho$ for $\sigma, \rho \in \mathrm{Sym}_k$ if we can obtain σ by applying a circular shift to ρ. This reduces the number of SDPs that we have to solve from $k!$ to $(k-1)!$. Furthermore, we note that we can solve these SDPs independently of each other. Therefore, we can use parallel computing to reduce the running time. We present some numerical results in Table 6.6. From this table we can see that our upper bound is very accurate and that we do not run out of memory as in Chapter 5.3.1. However, the running time grows more than exponentially in k.

5.3.3 Minimum rank solutions

It turns out that the upper bound from Chapter 5.3.2 is very close to the exact value of $1/4 c_{\mathrm{EHZ}}(C)$ for most of the two-dimensional random instances that we consider in Table 6.6. Therefore, we would like to verify whether the upper bound is indeed equal to the exact value. The following strategy is one way to achieve such a verification. First, we compute the upper bound

$$\max\{\widehat{p}_\sigma \colon \sigma \in \mathrm{Sym}_k\}$$

and we let $\sigma^* \in \mathrm{Sym}_k$ be a permutation for which the maximum is attained. In particular, we have

$$\frac{1}{4c_{\mathrm{EHZ}}(C)} = \max\{p^*_\sigma \colon \sigma \in \mathrm{Sym}_k\} \leqslant \max\{\widehat{p}_\sigma \colon \sigma \in \mathrm{Sym}_k\} = \widehat{p}_{\sigma^*}. \tag{5.30}$$

Assume that in the SDP (5.25) for σ^* there is an optimal solution X with $\operatorname{rk} X = 1$. Then we have $X = xx^T$ for some $x \in \mathbb{R}^k$. Furthermore, due to the constraint $X_{i,j} \geqslant 0$ for every $i, j \in \{1, \ldots, k\}$ we can assume that x is nonnegative. Thus, we find that X is completely positive and that it is a feasible solution of the CP (5.23) for σ^*. Therefore,

$$\widehat{p}_{\sigma^*} \leqslant p^*_{\sigma^*}. \tag{5.31}$$

On the other hand, the SDP is a relaxation of the CP. This means that (5.31) holds with equality. It follows that the first maximum in (5.30) is attained for σ^* and

$$\frac{1}{4c_{\mathrm{EHZ}}(C)} = p^*_{\sigma^*} = \widehat{p}_{\sigma^*}.$$

Summerizing these considerations, we get that the existence of an optimal solution of (5.25) for σ^* with rank one implies that the bound

$$\frac{1}{4c_{\mathrm{EHZ}}(C)} \leqslant \max\{\widehat{p}_\sigma \colon \sigma \in \mathrm{Sym}_k\}$$

holds with equality. Motivated by this approach we give the following bound on $\operatorname{rk} X$, where X is a feasible solution of the SDP (5.25).

Theorem 5.3.8. *Let $C \subseteq \mathbb{R}^{2n}$ be a polytope with $0 \in \operatorname{int} C$ and let $v_1, \ldots, v_k$ be the vertices of C°. Furthermore, let $X \in \mathcal{S}^k_{\geq 0}$ be a feasible solution of the SDP* (5.25) *for some $\sigma \in \mathrm{Sym}_k$. Then:*

$$\operatorname{rk} X \leqslant k - 2n.$$

Proof. Recall that the matrix A in (5.25) is the $(2n \times k)$-matrix whose ith column is $v_{\sigma(i)}$ for every $i \in \{1, \ldots, k\}$. Using spectral decomposition of the positive semidefinite matrix X, we get

$$0 = \left\langle A^T A, X \right\rangle = \sum_{i=1}^{k} \lambda_i \left\langle A^T A, u_i u_i^T \right\rangle = \sum_{i=1}^{k} \lambda_i u_i^T A^T A u_i = \sum_{i=1}^{k} \lambda_i ||Au_i||^2,$$

where $u_1, \ldots, u_k \in \mathbb{R}^k$ are the orthonormal eigenvectors and $\lambda_1, \ldots, \lambda_k \geqslant 0$ are the eigenvalues of X. Since every term in the last sum is nonnegative, it follows that for every $i \in \{1, \ldots, k\}$ we have $\lambda_i = 0$ or $Au_i = 0$. Hence, every eigenvector corresponding to a positive eigenvalue of X is contained in $\ker A$. Next, we observe that C is full-dimensional because it has nonempty interior. Therefore, there are $2n$ linearly independent vectors among $v_1, \ldots, v_k$. This implies $\operatorname{rank} A = 2n$ and by the rank-nullity theorem

$$\ker A = k - 2n.$$

So, X has at most $k - 2n$ eigenvectors that correspond to positive eigenvalues. If we count eigenvalues according to their multiplicity, this means that X has at most $k - 2n$ nonzero eigenvalues which implies the claim. □

An immediate consequence of Theorem 5.3.8 is that the upper bound from Chapter 5.3.2 is exactly $1/4c_{\mathrm{EHZ}}(C)$ if C is a simplex. The reason for this is that a nonempty polytope $C \subseteq \mathbb{R}^{2n}$ is a simplex if and only if it has $2n+1$ facets. As we note in (5.3) the vertices of C° correspond to facets of C. Thus, Theorem 5.3.8 states that every feasible solution of the SDP (5.25) for arbitrary $\sigma \in \mathrm{Sym}_k$ has rank at most one. In particular, the same is true for every optimal solution.

If C is not a simplex, we can use an SDP technique to estimate the minimal rank among all optimal solutions of an SDP. There are several methods to solve such rank minimization problems. We present the one that is given in [85]. For this, we define the *spectral norm*

$$||X|| := \sigma_1(X)$$

and the *nuclear norm*

$$||X||_* := \sum_{i=1}^{\min\{r,s\}} \sigma_i(X)$$

for every $X \in \mathbb{R}^{r\times s}$. Here,

$$\sigma_1(X) \geqslant \ldots \geqslant \sigma_{\min\{r,s\}}(X) \geqslant 0$$

denote the singular values of X. We recall that the ith largest singular value of X is the square root of the ith largest eigenvalue of X^TX for every $i \in \{1,\ldots,\min\{r,s\}\}$. The idea is now to formulate the nuclear norm $||X||_*$ as an SDP. Then we can add constraints to this SDP and obtain a way to find an optimal solution for the SDP (5.25) that has minimal nuclear norm. In practice, this is a good heuristic for an optimal solution with minimal rank. First, we present a statement that relates the spectral norm and the nuclear norm to each other. The reason is that there is an easy representation of the spectral norm as an SDP. We exploit this fact in the proof of the following theorem to deduce an SDP reformulation of the nuclear norm as well.

Theorem 5.3.9. *[85] The nuclear norm is the dual norm of the spectral norm, i.e.:*

$$||X||_* = \max\left\{\langle X, Y\rangle : Y \in \mathbb{R}^{r\times s},\ ||Y|| \leqslant 1\right\}$$

for all $X \in \mathbb{R}^{r\times s}$.

Proof. Let $Z \in \mathbb{R}^{r\times s}$ and let $t \in \mathbb{R}$. The matrix Z^TZ is positive semidefinite and we can write Z^TZ using spectral decomposition, i.e. $Z^TZ = P^T\Sigma P$, where P is an orthogonal matrix and Σ is a diagonal matrix whose diagonal consists of the eigenvalues of Z^TZ. In particular, we have

$$t^2I - Z^TZ = t^2P^TP - P^T\Sigma P = P^T(t^2I - \Sigma)P.$$

Therefore, we have that $t^2I - Z^TZ$ is positive semidefinite if and only if the largest eigenvalue of Z^TZ is at most t^2. In other words:

$$||Z|| \leqslant t \iff t^2I - Z^TZ \succeq 0 \iff \begin{pmatrix} tI & Z \\ Z^T & tI \end{pmatrix} \succeq 0. \tag{5.32}$$

The last equivalence follows from Theorem 3.2.7 for $t \neq 0$. It is easy to see that $t = 0$ implies $Z = 0$ in all three statements. Thus, we can rewrite the spectral norm as an SDP:

$$||Z|| = \min\left\{t \in \mathbb{R}\colon \begin{pmatrix} tI & Z \\ Z^T & tI \end{pmatrix} \succeq 0\right\}.$$

Now we let $X \in \mathbb{R}^{r\times s}$ be a matrix. X has a singular value decomposition [54] which means that there are orthogonal matrices $U \in O(r), V \in O(s)$ and a matrix $\Sigma \in \mathbb{R}^{r\times s}$ such that

$$X = U\Sigma V^T.$$

Moreover, the matrix Σ is given by $\Sigma_{i,i} = \sigma_i(X)$ and $\Sigma_{i,j} = 0$ for all $i \in \{1,\dots,r\}$ and $j \in \{1,\dots,s\}$ with $i \neq j$. We let $\widehat{Y} := UV^T$. We can immediately see the singular value decomposition of $\widehat{Y}$. This decomposition yields that every singular value of $\widehat{Y}$ is one. In particular we have $||\widehat{Y}|| = 1$ and

$$\left\langle X, \widehat{Y}\right\rangle = \operatorname{tr}(X^T\widehat{Y}) = \operatorname{tr}(V\Sigma U^T UV^T) = \operatorname{tr}(V^T V\Sigma U^T U) = \operatorname{tr}\Sigma = \sum_{i=1}^{\min\{r,s\}} \sigma_i(X).$$

Thus, we get

$$||X||_* = \left\langle X, \widehat{Y}\right\rangle \leqslant \max\left\{\langle X, Y\rangle : Y \in \mathbb{R}^{r\times s},\ ||Y|| \leqslant 1\right\}. \tag{5.33}$$

For the converse we use (5.32) to rewrite the dual norm of the spectral norm as an SDP:

$$\begin{aligned} \max\left\{\langle X, Y\rangle : Y \in \mathbb{R}^{r\times s},\ ||Y|| \leqslant 1\right\} \ = \ & \max \quad \langle X, Y\rangle \\ & \text{s. t.} \quad Y \in \mathbb{R}^{r\times s} \\ & \qquad \begin{pmatrix} I & Y \\ Y^T & I \end{pmatrix} \succeq 0. \end{aligned}$$

We construct the corresponding dual SDP and use weak duality (Theorem 3.2.3 (i)) to get

$$\begin{aligned} \max\left\{\langle X, Y\rangle : Y \in \mathbb{R}^{r\times s},\ ||Y|| \leqslant 1\right\} \ \leqslant \ & \min \quad \frac{1}{2}\left(\langle I, Y_1\rangle + \langle I, Y_2\rangle\right) \\ & \text{s. t.} \quad Y_1 \in \mathbb{R}^{r\times r},\ Y_2 \in \mathbb{R}^{s\times s} \\ & \qquad \begin{pmatrix} Y_1 & -X \\ -X^T & Y_2 \end{pmatrix} \succeq 0 \end{aligned} \tag{5.34}$$

Next, we construct a feasible solution of the dual SDP in (5.34). For this we write the singular value decomposition of X as

$$X = U'\Sigma'(V')^T$$

with $U' \in \mathbb{R}^{r\times\ell}$, $V' \in \mathbb{R}^{s\times\ell}$ and $\Sigma' \in \mathbb{R}^{\ell\times\ell}$, where $\ell = \min\{r, s\}$. We achieve this decomposition by removing zero columns and rows from Σ and by removing columns from

U and V accordingly. We let $Y_1 = U'\Sigma'(U')^T$ and $Y_2 = V'\Sigma'(V')^T$. It is easy to verify that (Y_1, Y_2) is a feasible solution for the dual SDP in (5.34):

$$\begin{pmatrix} Y_1 & -X \\ -X^T & Y_2 \end{pmatrix} = \begin{pmatrix} U'\Sigma'(U')^T & -U'\Sigma'(V')^T \\ -V'\Sigma'(U')^T & V'\Sigma'(V')^T \end{pmatrix} = \begin{pmatrix} U' \\ -V' \end{pmatrix} \Sigma' \begin{pmatrix} U' \\ -V' \end{pmatrix}^T \succeq 0.$$

Note that Σ' is a square diagonal matrix with nonnegative entries. The columns of U' are orthonormal because $U \in O(r)$. Therefore, we have

$$\langle I, Y_1 \rangle = \operatorname{tr} Y_1 = \operatorname{tr}\left(U'\Sigma'(U')^T\right) = \operatorname{tr}\left((U')^T U'\Sigma'\right) = \operatorname{tr}\Sigma' = \sum_{i=1}^{\min\{r,s\}} \sigma_i(X).$$

The columns of V' are orthonormal as well and we can make a similar calculation for $\langle I, Y_2 \rangle$. Together, we find

$$\frac{1}{2}\left(\langle I, Y_1 \rangle + \langle I, Y_2 \rangle\right) = ||X||_*.$$

(5.34) now implies

$$\max\left\{\langle X, Y \rangle : Y \in \mathbb{R}^{r\times s},\ ||Y|| \leqslant 1\right\} \leqslant ||X||_*$$

and the theorem follows with (5.33). □

An important feature of the proof of Theorem 5.3.9 is that (5.34) holds with equality. Thus, we can express the nuclear norm as an SDP. As mentioned earlier, this is useful if we want to examine the rank of some matrices because the nuclear norm is in a certain sense the best convex approximation of the rank. More precisely, we have the following theorem that is due to Fazel [36].

Theorem 5.3.10. *[36] Let* $M = \{X \in \mathbb{R}^{r\times s} : ||X|| \leqslant 1\}$. *The nuclear norm is a convex function on* M *and*

$$||X||_* \leqslant \operatorname{rk} X$$

for every $X \in M$. *Moreover, if* $f : M \to \mathbb{R}$ *is a convex function with* $f(X) \leqslant \operatorname{rk} X$ *for every* $X \in M$, *then*

$$f(X) \leqslant ||X||_*$$

for every $X \in M$.

We point out that the first statement in Theorem 5.3.10, i.e.:

$$||X||_* \leqslant \operatorname{rk} X \tag{5.35}$$

for every matrix $X \in M$, is easy to prove. If $X \in M$, then the largest singular value of X is at most one. Thus, the nuclear norm

$$||X||_* = \sum_{i=1}^{\min\{r,s\}} \sigma_i(X)$$

is bounded by the number of nonzero singular values of X. The inequality (5.35) now follows from the fact that the rank of X is equal to the number of its nonzero singular values [54].

A more compact way to express Theorem 5.3.10 is to say that the nuclear norm is the *convex envelope* of the rank function [101]. Theorem 5.3.10 suggests that if we are interested in an optimal solution of the SDP (5.25) with minimal rank, then computing an optimal solution with minimal nuclear norm is a reasonable heuristic. Let us make this approach more precise. We assume we solved the SDP (5.25) for every $\sigma \in \mathrm{Sym}_k/\{\text{circular shifts}\}$. Let σ^* be a permutation for which the maximum

$$\max\{\widehat{p}_\sigma \colon \sigma \in \mathrm{Sym}_k\}$$

is attained. Then we solve the following SDP:

$$\begin{aligned} \min \ & \frac{1}{2}\left(\langle I, Y_1\rangle + \langle I, Y_2\rangle\right) \\ \text{s. t. } & Y_1, Y_2 \in \mathbb{R}^{k\times k},\ X \in \mathcal{S}^k_{\geq 0} \\ & \begin{pmatrix} Y_1 & -X \\ -X & Y_2 \end{pmatrix} \succeq 0 \\ & \langle Q, X\rangle = \widehat{p}_{\sigma^*} \\ & \left\langle A^T A, X\right\rangle = 0 \\ & \left\langle ee^T, X\right\rangle = 1 \\ & X_{i,j} \geqslant 0,\ \forall\, i,j \in \{1,\dots,k\}, \end{aligned} \tag{5.36}$$

where the ith column of the $(2n \times k)$-matrix A is $v_{\sigma^*(i)}$ and the matrix Q is given by

$$Q_{i,j} = \begin{cases} -\frac{1}{2} v_{\sigma^*(i)} J v_{\sigma^*(j)}, & \text{if } i > j, \\ 0, & \text{if } i \leqslant j, \end{cases}$$

for every $i, j \in \{1,\dots,k\}$. It is obvious that if (Y_1, Y_2, X) is a feasible solution for (5.36), then X is an optimal solution for (5.25). Thus, if we find that $\operatorname{rk} X = 1$ it follows that $\widehat{p}_{\sigma^*} = 1/4c_{\mathrm{EHZ}}(C)$ as we describe earlier in this chapter.

We note that to solve (5.36) it is crucial that we compute $\widehat{p}_{\sigma^*}$ accurately because there is no feasible matrix X that satisfies

$$\langle Q, X\rangle = \widehat{p}_{\sigma^*} + \varepsilon$$

for $\varepsilon > 0$. Therefore, we solve the SDP (5.25) for σ^* again with higher accuracy to obtain $\widehat{p}_{\sigma^*}$. For this, we use the SDPA-GMP solver [77], [78], [102], which is a slower solver that is based on SDPA and that is designed to compute highly accurate solutions. We use an accuracy of 10^{-10}, i.e. we compute a feasible matrix X^* such that

$$\left|\langle Q, X^*\rangle - \widehat{p}_{\sigma^*}\right| \leqslant 10^{-10}.$$

This is sufficient to ensure that the SDP (5.36) can be solved computationally if we plug in $\langle Q, X^*\rangle$ for $\widehat{p}_{\sigma^*}$.

In the last column of Table 6.6 we provide the numerical results for the same instances that we considered in Chapter 5.3.2. We see that our upper bound from Chapter 5.3.2 is often equal to the exact value of $1/4c_{\mathrm{EHZ}}(C)$ up to solver accuracy.

5.4 Lower bounds

Another way to estimate the quality of the upper bound that we describe in Chapter 5.3.2 is to find good lower bounds on $1/4c_{\mathrm{EHZ}}(C)$. To this end, we consider the maximization problem (5.4). We see that every feasible solution $(y_1, \dots, y_k, \sigma)$ yields a lower bound on the optimal value and hence also on $1/4c_{\mathrm{EHZ}}(C)$. Thus, the goal is to find a feasible solution that ideally has an objective value close to the optimal value. We get a decent candidate for the permutation σ by computing the upper bound from Chapter 5.3.2. More precisely, we take the permutation σ^* for which the maximum

$$\max\{\widehat{p}_\sigma \colon \sigma \in \mathrm{Sym}_k\}$$

is attained. To find reasonable candidates for $y_1, \dots, y_k$ we consider the corresponding quadratic maximization problem (5.22) and compute a local optimum. For this, we use the TRUST-CONSTR method from SciPy [59]. This method relies on [27] and employs sequential quadratic programming techniques combined with a trust region approach. This means that we solve (5.22) iteratively, similar to a Newton method. In each step the objective function and the constraints are made more manageable by approximating them by quadratic or linear models. At the same time the search of the next iterate is restricted to a subregion such that the quadratic or linear models are good approximations on this subregion. The input for this method consists of the constraints of (5.22) and the objective function

$$f \colon \mathbb{R}^k \to \mathbb{R}, \;\; f(y) = -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{i-1} y_i y_j v_{\sigma^*(i)}^T J v_{\sigma^*(j)}$$

as well as the gradient and Hessian of the objective function:

$$(\nabla f(y))_i = -\frac{1}{2} \left(\sum_{j=1}^{i-1} y_j v_{\sigma^*(i)}^T J v_{\sigma^*(j)} + \sum_{j=i+1}^{k} y_j v_{\sigma^*(j)}^T J v_{\sigma^*(i)} \right),$$

$$(H_f(y))_{i,j} = \begin{cases} 0, & i = j, \\ -\frac{1}{2} v_{\sigma^*(i)}^T J v_{\sigma^*(j)}, & i > j, \\ -\frac{1}{2} v_{\sigma^*(j)}^T J v_{\sigma^*(i)}, & i < j, \end{cases}$$

for all $i, j \in \{1, \dots, k\}$. Additionally, the TRUST-CONSTR method requires a feasible solution of (5.22) as a starting point. More precisely, we need to input a vector $y_0 \in \mathbb{R}^k_{\geqslant 0}$ such that $By = b$, where

$$B = \begin{pmatrix} v_{\sigma^*(1)} & \cdots & v_{\sigma^*(k)} \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{(2n+1)\times k}, \;\; b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^{2n+1}.$$

One way to find such a feasible starting point is to employ an LP technique. More precisely, we let $c \in \mathbb{R}^k$ and we pick y_0 as an optimal solution of the following LP:

$$\begin{aligned} \max\ & c^T y \\ & y \in \mathbb{R}^k_{\geqslant 0} \\ & By = b. \end{aligned}$$

Note that this maximum is finite for every vector c because the set of feasible solutions is contained in the standard simplex

$$\Delta_{k-1} := \{y \in \mathbb{R}^k \colon y_1 + \ldots + y_k = 1,\ y \geqslant 0\}$$

and hence is bounded.

With this approach we get a feasible starting point that lies on the boundary of the set of feasible solutions of (5.22). Therefore, if this boundary contains many saddle points of the objective function, we might find a feasible starting point for which the TRUST-CONSTR method terminates immediately without having found a local optimum. Next, we discuss an approach that is capable of producing a feasible starting point in the interior of the set of feasible solutions.

First, we choose a uniformly distributed random vector in $[0,1]^k$. Then we project this vector onto the hyperplane $\{x \colon Bx = b\}$ and check whether its entries are still nonnegative after this projection. If this is the case, then this point is a feasible solution of (5.22). Otherwise, we repeat this process until we found a feasible solution.

Algorithm 4 Feasible starting point

while no y_0 found so far **do**
  Generate a random vector u_0 uniformly distributed in $[0,1]^k$.
  Let B_1^T be the first row of B. Set

$$u_1 := u_0 + \frac{b_1 - B_1^T u_0}{||B_1||^2} B_1.$$

  for $i \in \{2, \ldots, 2n+1\}$ **do**
    Let B_j^T be the jth row of B for every $j \in \{1, \ldots, 2n+1\}$. Find $w_i \in \mathbb{R}^k$ such that

$$\begin{pmatrix} B_1^T \\ \vdots \\ B_{i-1}^T \\ B_i^T \end{pmatrix} w_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

    Set $u_i := u_{i-1} + (b_i - B_i^T u_{i-1}) w_i$.
  end for
  if $u_{2n+1} \geqslant 0$ **then**
    Output $y_0 := u_{2n+1}$.
  end if
end while

If y_0 is the output from Algorithm 4, then it is easy to see that $By_0 = b$ holds. Due to the third line of the algorithm we have

$$B_1^T u_1 = B_1^T u_0 + \frac{b_1 - B_1^T u_0}{||B_1||^2} B_1^T B_1 = b_1.$$

Furthermore, for every $i, j \in \{1, \ldots, 2n+1\}$ with $j < i$ the lines 5 and 6 imply

$$B_i^T u_i = B_i^T u_{i-1} + (b_i - B_i^T u_{i-1}) B_i^T w_i = b_i,$$
$$B_j^T u_i = B_j^T u_{i-1} + (b_i - B_i^T u_{i-1}) B_j^T w_i = B_j^T u_{i-1}.$$

From this we conclude that for every $i \in \{1, \ldots, 2n+1\}$ the following holds:

$$B_i^T y_0 = B_i^T u_{2n+1} = B_i^T u_{2n} = \ldots = B_i^T u_i = b_i.$$

It remains to show that in line 5 it is always possible to find a vector $w_i \in \mathbb{R}^k$.

Lemma 5.4.1. *Let $C \subseteq \mathbb{R}^{2n}$ be a polytope with $0 \in \operatorname{int} C$ and let $v_1, \ldots, v_k$ be the vertices of C°. Furthermore, let $B \in \mathbb{R}^{(2n+1)\times k}$ be defined by*

$$B = \begin{pmatrix} v_1 & \cdots & v_k \\ 1 & \cdots & 1 \end{pmatrix}.$$

Then for every $i \in \{2, \ldots, 2n+1\}$ there is $w_i \in \mathbb{R}^k$ such that:

$$\begin{pmatrix} B_1^T \\ \vdots \\ B_{i-1}^T \\ B_i^T \end{pmatrix} w_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Proof. To simplify the notation we fix $i \in \{2, \ldots, 2n+1\}$ and let

$$\widehat{B} = \begin{pmatrix} B_1^T \\ \vdots \\ B_{i-1}^T \\ B_i^T \end{pmatrix} \in \mathbb{R}^{i\times k}, \quad \widehat{b} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^i.$$

We prove the statement by contradiction, so assume it is false. Then there are no vectors $w_i^+, w_i^- \in \mathbb{R}^k_{\geqslant 0}$ such that

$$\widehat{b} = \widehat{B}(w_i^+ - w_i^-) = \widehat{B}w_i^+ - \widehat{B}w_i^- = \begin{pmatrix} \widehat{B} & -\widehat{B} \end{pmatrix} \begin{pmatrix} w_i^+ \\ w_i^- \end{pmatrix}.$$

If we apply Farkas' Lemma [89], we find that there is a vector $z \in \mathbb{R}^i$ such that

$$z^T \begin{pmatrix} \widehat{B} & -\widehat{B} \end{pmatrix} \geqslant 0 \text{ and } z^T \widehat{b} < 0.$$

On the one hand, this implies $z_i < 0$ and in particular $z \neq 0$. On the other hand, we get $z^T \widehat{B} = 0$. This means that the rows of $\widehat{B}$ and hence also the rows of B are linearly dependent. More precisely,

$$y^T B = y^T \begin{pmatrix} v_1 & \cdots & v_k \\ 1 & \cdots & 1 \end{pmatrix} = 0$$

has a solution $y \neq 0$. It follows that $v_1, \ldots, v_k$ lie in the hyperplane

$$H = \{x \in \mathbb{R}^{2n} \colon x^T y = -y_{2n+1}\}.$$

In particular, $C^\circ = \operatorname{conv}\{v_1, \ldots, v_k\}$ is contained in H and thus has no interior in $\mathbb{R}^{2n}$. But from (5.17) we see that $0 \in \operatorname{int} C^\circ$. So, the lemma follows by contradiction. □

In Table 6.7 we compare our lower bound with the upper bound from Chapter 5.3.2. We carried out the computation of the lower bound on a Dell OptiPlex 9020 MT Desktop-PC with Intel Core i7-4770 processor, 3.4 GHz (capable of running 8 threads).

Chapter Six

Outlook

In this thesis we provide several ways to compute bounds on the EHZ capacity. One may wonder how these bounds can be improved or what other approaches appear promising. To conclude this thesis, we suggest some ideas in this direction.

Again, we consider the maximization problem (5.4). As we can see from the results of Chapter 5 we have a decent way to approximate an optimal solution $(y_1^*, \dots, y_k^*)$, if we are given a permutation σ^* for which the maximum is achieved. However, finding such a permutation remains a difficult problem. We address this problem in Chapter 5.2 using an eigenvalue based QAP technique. One can also consider other approaches to handle the QAP that we face in Chapter 5.2. Two general techniques that come to mind in this context are linear and semidefinite relaxations of QAP (see for instance [86] and [104]). Recall that in Chapter 5.2 the goal is to solve the maximization problem

$$\max\left\{-\frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{h-1} z_{\sigma(h)}z_{\sigma(j)}v_{\sigma(h)}^T J v_{\sigma(j)} : z \in M,\ \sigma \in \mathrm{Sym}_k\right\},$$

$$M = \left\{z \in \mathbb{R}^k_{\geqslant 0} \colon \sum_{j=1}^{k} z_j v_j = 0,\ \sum_{j=1}^{k} z_j = 1\right\},$$

where $v_1, \dots, v_k$ are the vertices of C° for some polytope $C \in \mathbb{R}^{2n}$ with $0 \in \operatorname{int} C$. The idea is now to fix $z \in M$ and to consider an LP or SDP that yields an upper bound on

$$\max\left\{-\frac{1}{2}\sum_{h=1}^{k}\sum_{j=1}^{h-1} z_{\sigma(h)}z_{\sigma(j)}v_{\sigma(h)}^T J v_{\sigma(j)} : \sigma \in \mathrm{Sym}_k\right\}.$$

Then, in a second step, one could aim to incorporate the optimization over M into this LP or SDP.

Problem 1. *Employ different QAP techniques to compute bounds on the optimal value of the maximization problem* (5.5) *(and hence on* $1/4c_{\mathrm{EHZ}}(C)$*).*

Aside from techniques with which one can compute bounds on the solution of a QAP, there are also strategies that aim to solve a QAP exactly. Branch and bound algorithms are an example for such exact solution algorithms. These algorithms should also be considered to handle the QAP in Chapter 5.2.

Problem 2. *Employ exact solution algorithms to determine the optimal value of the maximization problem* (5.5) *(and hence* $1/4c_{\mathrm{EHZ}}(C)$*).*

Our approach to compute bounds on $1/4c_{\mathrm{EHZ}}(C)$ via completely positive optimization yields two additional problems that are left open. First, we observe from our numerical results that the bound discussed in Chapter 5.3.2 (see Table 6.6) is often equal to the exact value of $1/4c_{\mathrm{EHZ}}(C)$ because we are able to find optimal matrices with rank one. This raises the following problem.

Problem 3. *For a polytope* $C \subseteq \mathbb{R}^{2n}$ *with* $0 \in \operatorname{int} C$ *consider the SDP* (5.25) *with optimal value* $\widehat{p}_{\sigma *}$*, where* $\sigma^* \in \mathrm{Sym}_k$ *is a permutation such that*

$$\widehat{p}_{\sigma *} = \max\{\widehat{p}_{\sigma} \colon \sigma \in \mathrm{Sym}_k\}.$$

For which polytopes C does this SDP have an optimal solution with rank one?

Second, we note that the bound from Chapter 5.3.2 provides better bounds than the one given in Chapter 5.3.1 (see Table 6.5). In fact, if we let $v_1, \dots, v_k$ be the vertices of C°, then

$$\begin{aligned}
&\max_{y_1,\dots,y_k,\sigma} \; -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} y_i y_j v_{\sigma(i)}^T J v_{\sigma(j)} \\
&\text{s. t.} \;\; \sum_{i=1}^{k} y_i v_{\sigma(i)} = 0 \\
&\qquad \sum_{i=1}^{k} y_i = 1 \\
&\qquad y_i \geqslant 0 \;\; \forall\, i \in \{1,\dots,k\} \\
&\qquad \sigma \in \mathrm{Sym}_k
\end{aligned}
\qquad \leqslant \qquad
\begin{aligned}
&\max_{\substack{y_1,\dots,y_k,\\ w_1,\dots,w_k}} \; -\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{i-1} y_i y_j w_i^T J w_j \\
&\text{s. t.} \;\; \sum_{i=1}^{k} y_i w_i = 0 \\
&\qquad \sum_{i=1}^{k} y_i = 1 \\
&\qquad y_i \geqslant 0 \;\; \forall\, i \in \{1,\dots,k\} \\
&\qquad w_1,\dots,w_k \in C^\circ
\end{aligned}
\tag{6.1}$$

because every feasible solution of the problem on the left-hand side yields a feasible solution of the problem on the right-hand side. Note that we consider these quadratic programs in Chapter 5.3.1 and Chapter 5.3.2. By Theorem 5.3.5 the inequality in (6.1) still holds if we replace each of these two quadratic optimization problems with a suitable CP. However, the question remains whether the inequality also holds if we apply semidefinite relaxation to both CPs. Our numerical results suggest that the answer is positive but the proof seems to be cumbersome since we consider large coefficient matrices in the CP reformulation of the problem on the right-hand side of (6.1).

Problem 4. *Prove that the bound on* $1/4c_{\mathrm{EHZ}}(C)$ *given in Chapter 5.3.2 (see* (5.26)*) is at least as good as the one given in Chapter 5.3.1 (see* (5.21)*) for every polytope* $C \in \mathbb{R}^{2n}$ *with* $0 \in \operatorname{int} C$*.*

For another idea for future research we recall Viterbo's conjecture which we state in Chapter 2.4.2.

Conjecture 6.0.1. *[99] Let c be a symplectic capacity and let* $C \subseteq \mathbb{R}^{2n}$ *be a convex set. Then:*

$$\frac{c(C,\omega_0)}{c(B_1(0),\omega_0)} \leqslant \left(\frac{\operatorname{vol} C}{\operatorname{vol} B_1(0)}\right)^{1/n}.$$

As stated earlier, we get a particularly interesting case of this conjecture if we fix $c = c_{\mathrm{EHZ}}$ and consider convex sets of the form $C = K \times K^\circ$, where K is a centrally symmetric, convex set. The reason for this is that this special case is equivalent to the prominent symmetric Mahler conjecture. In Chapter 4 we provide an algorithm that computes $c_{\mathrm{EHZ}}(K \times T)$ for two polytopes K and T. We can use this algorithm to test Viterbo's conjecture. More precisely, if there is a centrally symmetric polytope K such that Viterbo's conjecture is false for $C = K \times K^\circ$, then we can use our algorithm to verify

$$\frac{c_{\mathrm{EHZ}}(K \times K^\circ)}{c_{\mathrm{EHZ}}(B_1(0))} > \left(\frac{\operatorname{vol} K \times K^\circ}{\operatorname{vol} B_1(0)} \right)^{1/n}.$$

It remains to find such a counterexample, if there is one.

Problem 5. *Find a centrally symmetric polytope K such that*

$$\frac{c_{\mathrm{EHZ}}(K \times K^\circ)}{c_{\mathrm{EHZ}}(B_1(0))} > \left(\frac{\operatorname{vol} K \times K^\circ}{\operatorname{vol} B_1(0)} \right)^{1/n}.$$

Appendix

In this appendix we provide numerical results for the algorithms and methods in this thesis. In Table 6.1 and 6.2 we present results for the Algorithms 1 and 2. Table 6.3 and 6.4 treat the upper bound on $1/4c_{\mathrm{EHZ}}(C)$ that is based on a QAP approach (see Chapter 5.2). Table 6.5 provides the upper bound on $1/4c_{\mathrm{EHZ}}(C)$ that we describe in Chapter 5.3.1. Afterwards, we state the results from Chapter 5.3.2 and Chapter 5.3.3 in Table 6.6. Lastly, in Table 6.7, we compare the upper bounds from Table 6.6 with the lower bounds that we describe in Chapter 5.4.

As stated in Chapter 4 we generate random polytopes as input for Algorithms 1 and 2 by taking the convex hull of normally distributed random points. We do the same to generate the input for the methods in Chapter 5. For better comparability we use the same instances in the Tables 6.3, 6.5, 6.6 and 6.7. These instances are enumerated in the first column of each of these tables. A more detailed description on how to replicate this input is available on the website

`www.github.com/S-Krupp/EHZ-capacity-of-polytopes`

together with the implementation of our algorithms.

Aside from a polytope C, the methods in Chapter 5 require the polar set C° as an input. This means that the time to calculate C° is not included in the running time in the following tables.

In Chapter 2.3 we mention that for every symplectic capacity c and every compact, connected set $D \subseteq \mathbb{R}^2$ with smooth boundary, we have

$$c(D, \omega_0) = \operatorname{vol} D.$$

Together with Lemma 2.3.3 we have a way to determine the exact value of $1/4c_{\mathrm{EHZ}}(C)$ for every two-dimensional polytope C. Thus, we provide this exact value in the Tables 6.3, 6.4, 6.5, 6.6 and 6.7 if possible to demonstrate the quality of our bounds.

# facets	dim	time 2 bp.	time 3 bp.	time 4 bp.	time 5 bp.
10	2	0.00831	0.30746	-	-
15	2	0.00973	0.77826	-	-
20	2	0.01228	1.90130	-	-
25	2	0.01535	3.68928	-	-
30	2	0.01885	6.87248	-	-
35	2	0.02242	10.98917	-	-
40	2	0.02611	17.82252	-	-
45	2	0.02676	26.24746	-	-
50	2	0.03756	37.58889	-	-
60	2	0.04348	68.84613	-	-
70	2	0.07126	121.99939	-	-
80	2	0.08522	181.44979	-	-
90	2	0.09377	276.15113	-	-
100	2	0.12422	390.50119	-	-
110	2	0.14459	516.93047	-	-
120	2	0.16201	706.91467	-	-
130	2	0.16660	887.57707	-	-
140	2	0.23777	1145.28408	-	-
150	2	0.26617	1400.09367	-	-
14	3	0.00983	0.02658	2.78469	-
20	3	0.01315	0.05639	12.75729	-
24	3	0.01319	0.10031	25.36455	-
30	3	0.01869	0.16165	69.23203	-
34	3	0.01990	0.30546	121.10618	-
40	3	0.02877	0.43128	281.39158	-
44	3	0.02691	0.50671	456.25295	-
50	3	0.03799	0.76186	755.02158	-
54	3	0.04065	0.96421	1091.77615	-
60	3	0.04458	1.32361	1646.65092	-
64	3	0.04991	1.61337	2158.03637	-
70	3	0.07663	2.09306	2849.52804	-
11	4	0.00991	0.01654	0.05954	2.36219
15	4	0.00870	0.02263	0.17406	19.43698
20	4	0.01176	0.04806	0.56930	54.60817
25	4	0.01730	0.10708	1.58835	245.41436
30	4	0.02492	0.24002	4.56021	961.83634
35	4	0.02928	0.36108	9.92964	2171.25146
40	4	0.03232	0.50996	19.76087	4201.35654

Table 6.1: Running times of Algorithm 1. The billiard table K is a polytope. The first two columns contain the number of facets and the dimension of K. The last four columns contain the running time for 2, 3, 4 and 5 bouncing points in seconds.

$\lvert V(K)\rvert$	$\lvert V(T)\rvert$	time Algorithm 2	time Algorithm 3
5	5	1.28237	0.22158
10	10	7.85068	2.54632
15	15	27.39097	11.71706
20	20	54.75007	25.03621
25	25	82.30637	60.11260
30	30	125.05111	110.61238
35	35	170.22497	181.65273
40	40	259.88731	302.30844
45	45	266.73415	385.03827
50	50	361.56254	609.04153
55	55	451.56054	786.54793
5	10	3.02675	0.22164
5	15	5.57299	0.22637
5	20	11.40925	0.22114
5	25	16.91015	0.23931
5	30	19.89903	0.21554
5	35	23.96365	0.39383
5	40	29.05107	0.21106
5	45	32.18348	0.54072
5	50	36.41029	0.38885
5	55	49.02657	0.57020
5	65	59.80655	0.79811
5	75	67.33951	0.72834
10	5	3.39280	1.19793
15	5	5.74532	4.32675
20	5	10.50168	11.31948
25	5	14.59203	24.99738
30	5	17.76183	45.60183
35	5	20.62535	90.62127
40	5	23.89690	137.52914
45	5	25.73543	170.43779
50	5	30.23246	266.66650
55	5	33.68478	345.84228
65	5	41.49229	558.22820
75	5	51.92742	937.36931

Table 6.2: Running times of Algorithm 2 and Algorithm 3. All numbers are given in seconds.

no.	$2n$	k	upper bound	$1/4c_{\mathrm{EHZ}}(C)$	running time in s.
1	2	3	0.118312739	0.03943758	0.03
2	2	3	0.654845637	0.21828188	0.06
3	2	4	0.579574375	0.06553471	0.05
4	2	4	0.571432932	0.06040702	0.06
5	2	5	3.252296476	0.19042428	0.06
6	2	5	0.736995517	0.03465068	0.08
7	2	6	1.126261845	0.02836316	0.08
8	2	6	0.720733241	0.02526072	0.09
9	2	7	0.985491510	0.01962865	0.08
10	2	7	0.715791752	0.01797603	0.07
11	2	8	0.763852162	0.01298729	0.07
12	2	8	0.944722052	0.01523280	0.09
13	2	9	1.487177140	0.01671262	0.10
14	2	9	1.373066146	0.01324055	0.10
15	2	10	1.490686021	0.01303215	0.09
16	2	10	1.441573266	0.01467288	0.08
17	2	11	1.702583848	0.01286290	0.09
18	2	11	1.334970645	0.01073714	0.10
19	4	5	0.297548760	—	0.07
20	4	5	0.195625368	—	0.08
21	4	6	3.338312157	—	0.08
22	4	6	2.628991873	—	0.07
23	4	7	4.777343817	—	0.08
24	4	7	5.891492750	—	0.07
25	4	8	5.746260320	—	0.07
26	4	8	7.792411833	—	0.07
27	4	9	13.016826334	—	0.08
28	4	9	16.390665185	—	0.08
29	4	10	21.937612205	—	0.09
30	4	10	16.982943326	—	0.09
31	4	11	20.654944146	—	0.11
32	4	11	12.527661511	—	0.09
33	6	7	0.491114783	—	0.10
34	6	7	0.335022836	—	0.10
35	6	8	1.443657647	—	0.08
36	6	8	2.525658478	—	0.07
37	6	9	3.699497740	—	0.08
38	6	9	4.188805924	—	0.08
39	6	10	3.383507518	—	0.09
40	6	10	8.788380407	—	0.09

Table 6.3: Upper bounds based on the QAP approach in Chapter 5.2. For two-dimensional polytopes we provide the exact value of $1/4c_{\mathrm{EHZ}}(C)$.

no.	$2n$	k	upper bound	$1/4c_{\mathrm{EHZ}}(C)$	running time in s.
41	2	40	17.602496958	0.00909707	0.35
42	2	50	29.547775366	0.00898793	0.43
43	2	60	45.232497368	0.00895741	0.53
44	2	70	60.932335802	0.00890341	0.64
45	2	80	84.601556668	0.00890331	0.74
46	2	90	101.275455498	0.00888242	0.84
47	2	100	123.467778088	0.00888238	0.95
48	4	40	735.244449053	—	0.34
49	4	50	1062.545297909	—	0.45
50	4	60	1751.250513762	—	0.53
51	4	70	2655.205710218	—	0.63
52	4	80	3348.928392732	—	0.71
53	4	90	4302.098377014	—	0.83
54	4	100	4957.872036794	—	0.92
55	6	40	347.407734072	—	0.34
56	6	50	610.667190452	—	0.43

Table 6.4: Upper bounds based on the QAP approach in Chapter 5.2 for larger instances.

no.	$2n$	k	upper bound	$1/4c_{\mathrm{EHZ}}(C)$	running time in s.
1	2	3	0.039468058	0.03943758	0.13
2	2	3	0.218443732	0.218281879	0.15
3	2	4	0.069972860	0.065534708	0.57
4	2	4	0.065090727	0.060407024	0.58
5	2	5	0.266359869	0.190424277	2.19
6	2	5	0.042582561	0.034650682	1.50
7	2	6	0.039404677	0.028363161	5.58
8	2	6	0.034717004	0.025260717	5.51
9	2	7	0.030068372	0.019628645	16.24
10	2	7	0.029644262	0.017976027	15.67
11	2	8	0.025605698	0.012987287	34.90
12	2	8	0.027875957	0.015232795	57.22
13	2	9	0.033203063	0.016712622	118.96
14	2	9	0.026567933	0.013240549	134.52
15	2	10	0.026317105	0.013032145	408.47
16	2	10	0.029802683	0.014672875	393.38
17	2	11	0.028995359	0.012862896	956.24
18	2	11	0.024114858	0.010737144	1350.73
19	4	5	0.111109919	—	7.60
20	4	5	0.055562610	—	7.63
21	4	6	0.156993945	—	55.55
22	4	6	0.178815627	—	48.07
23	4	7	0.163989888	—	363.57
24	4	7	0.238847434	—	547.43
25	4	8	0.270707005	—	5962.98
26	4	8	0.189944156	—	2672.76
27	4	9	out of memory		
33	6	7	0.149188406	—	348.37
34	6	7	0.104485904	—	413.39
35	6	8	out of memory		

Table 6.5: Upper bounds by solving an SDP as described in Chapter 5.3.1 and the corresponding running time. Additionally, we provide the exact value of $1/4c_{\mathrm{EHZ}}(C)$ if the polytope is two-dimensional.

no.	$2n$	k	upper bound	$1/4c_{\mathrm{EHZ}}(C)$	running time in s.	rank heuristic
1	2	3	0.039438726	0.03943758	0.02	1
2	2	3	0.218282856	0.218281879	0.02	1
3	2	4	0.065536004	0.065534708	0.05	1
4	2	4	0.060408224	0.060407024	0.04	1
5	2	5	0.190435748	0.190424277	0.23	1
6	2	5	0.034654127	0.034650682	0.16	1
7	2	6	0.028366133	0.028363161	0.84	1
8	2	6	0.025263986	0.025260717	0.82	1
9	2	7	0.019632944	0.019628645	5.11	1
10	2	7	0.018180046	0.017976027	5.02	3
11	2	8	0.012992788	0.012987287	34.98	1
12	2	8	0.015237283	0.015232795	34.93	1
13	2	9	0.017762342	0.016712622	474.12	3
14	2	9	0.013244750	0.013240549	328.82	1
15	2	10	0.013035867	0.013032145	5768.32	1
16	2	10	0.015951592	0.014672875	6010.77	4
17	2	11	0.015016787	0.012862896	65693.18	5
18	2	11	0.011773859	0.010737144	66326.39	3
19	4	5	0.111035126	—	0.16	1
20	4	5	0.054042156	—	0.19	1
21	4	6	0.145718315	—	0.98	1
22	4	6	0.153409896	—	0.88	1
23	4	7	0.133809639	—	6.24	1
24	4	7	0.196341291	—	7.24	1
25	4	8	0.186138179	—	46.60	1
26	4	8	0.149438204	—	37.90	1
27	4	9	0.186649345	—	498.42	1
28	4	9	0.188494803	—	583.19	1
29	4	10	0.283382067	—	8988.94	1
30	4	10	0.181076891	—	5299.46	1
31	4	11	0.251035638	—	77898.48	4
32	4	11	0.200704143	—	90035.55	1
33	6	7	0.144871774	—	5.56	1
34	6	7	0.096864958	—	6.77	1
35	6	8	0.069709911	—	46.48	1
36	6	8	0.141613154	—	42.00	1
37	6	9	0.140578777	—	544.97	1
38	6	9	0.138142153	—	454.95	1
39	6	10	0.086869180	—	8206.18	1
40	6	10	0.129523513	—	8977.53	1

Table 6.6: Upper bounds by solving $(k-1)!$ SDPs as described in Chapter 5.3.2 and the corresponding running time. Additionally, we provide the exact value of $1/4c_{\mathrm{EHZ}}(C)$ if the polytope is two-dimensional as well as a heuristic for the minimal rank of an optimal solution as described in Chapter 5.3.3. If this heuristic is equal to one, then the upper bound is exact (up to solver accuracy).

no.	$2n$	k	lower bound	upper bound	$1/4c_{\mathrm{EHZ}}(C)$
1	2	3	0.039437580	0.039438726	0.03943758
2	2	3	0.218281879	0.218282856	0.218281879
3	2	4	0.065534708	0.065536004	0.065534708
4	2	4	0.060407024	0.060408224	0.060407024
5	2	5	0.190424277	0.190435748	0.190424277
6	2	5	0.034650682	0.034654127	0.034650682
7	2	6	0.028363161	0.028366133	0.028363161
8	2	6	0.025260717	0.025263986	0.025260717
9	2	7	0.019628643	0.019632944	0.019628645
10	2	7	0.013454116	0.018180046	0.017976027
11	2	8	0.012987287	0.012992788	0.012987287
12	2	8	0.015232794	0.015237283	0.015232795
13	2	9	0.014746672	0.017762342	0.016712622
14	2	9	0.013240447	0.013244750	0.013240549
15	2	10	0.013032123	0.013035867	0.013032145
16	2	10	0.012411318	0.015951592	0.014672875
17	2	11	0.012565752	0.015016787	0.012862896
18	2	11	0.010026363	0.011773859	0.010737144
19	4	5	0.111033394	0.111035126	—
20	4	5	0.054040204	0.054042156	—
21	4	6	0.145558669	0.145718315	—
22	4	6	0.153397393	0.153409896	—
23	4	7	0.133793204	0.133809639	—
24	4	7	0.196316775	0.196341291	—
25	4	8	0.186132486	0.186138179	—
26	4	8	0.149407543	0.149438204	—
27	4	9	0.186568474	0.186649345	—
28	4	9	0.188470827	0.188494803	—
29	4	10	0.283324215	0.283382067	—
30	4	10	0.180856019	0.181076891	—
31	4	11	0.208349487	0.251035638	—
32	4	11	0.200589322	0.200704143	—
33	6	7	0.144853224	0.144871774	—
34	6	7	0.096860887	0.096864958	—
35	6	8	0.069704162	0.069709911	—
36	6	8	0.141587402	0.141613154	—
37	6	9	0.140540242	0.140578777	—
38	6	9	0.138093848	0.138142153	—
39	6	10	0.086833645	0.086869180	—
40	6	10	0.129474486	0.129523513	—

Table 6.7: Lower bound from Chapter 5.4 and upper bound from Chapter 5.3.2 on $1/4c_{\mathrm{EHZ}}(C)$. Additionally, we provide the exact value if the polytope is two-dimensional.

Bibliography

[1] A. Abbondandolo and P. Majer. A non-squeezing theorem for convex symplectic images of the Hilbert ball. *Calculus of Variations and Partial Differential Equations*, 54:1469–1506, 2015.

[2] A. Akopyan and A. Balitskiy. Billiards in convex bodies with acute angles. *Israel Journal of Mathematics*, 216, 2015.

[3] A. Akopyan, A. Balitskiy, R. Karasev, and A. Sharipova. Elementary approach to closed billiard trajectories in asymmetric normed spaces. *Proceedings of the American Mathematical Society*, 144(10):4501–4513, 2016.

[4] A. Akopyan and R. Karasev. When different norms lead to same billiard trajectories? *European Journal of Mathematics*, 2020.

[5] A. Akopyan, R. Schwartz, and S. Tabachnikov. Billiards in ellipses revisited. *arXiv:2001.02934 [math.MG]*, 2020.

[6] N. Alkoumi and F. Schlenk. Shortest closed billiard orbits on convex tables. *Manuscripta Mathematica*, 147:365–380, 2014.

[7] C. Antony. Hofer-Zehnder capacity and its applications. *ETH Zürich, A semesterpaper submitted for the degree of MSc Mathematics*, 2013.

[8] MOSEK ApS. *MOSEK Optimizer API for Python 9.2.29*. `https://docs.mosek.com/9.2/pythonapi/index.html`, 2019.

[9] S. Artstein-Avidan, D. I. Florentin, Y. Ostrover, and D. Rosen. Duality of caustics in Minkowski billiards. *Nonlinearity*, 31(4):1197–1226, 2018.

[10] S. Artstein-Avidan, R. Karasev, and Y. Ostrover. From symplectic measurements to the Mahler conjecture. *Duke Math. J.*, 163(11):2003–2022, 2014.

[11] S. Artstein-Avidan and Y. Ostrover. Bounds for Minkowski billiard trajectories in convex bodies. *International Mathematics Research Notices*, 2014(1):165–193, 2012.

[12] Arvind, B. Dutta, N. Mukunda, and R. Simon. The real symplectic groups in quantum mechanics and optics. *Pramana*, 45(6):471–497, 1995.

[13] M. F. Atiyah. Convexity and commuting Hamiltonians. *Bulletin of the London Mathematical Society*, 14(1):1–15, 1982.

[14] S. Badrloo and A. Husseinzadeh Kashan. Combinatorial optimization of permutation-based quadratic assignment problem using optics inspired optimization. *Journal of Applied Research on Industrial Engineering*, 6(4):314–332, 2019.

[15] R. G. Bartle and D. R. Sherbert. *Introduction to Real Analysis, 4th Edition*. John Wiley & Sons, Incorporated, 2011.

[16] A. Barvinok. *A Course in Convexity*, volume 54 of *Graduate studies in mathematics*. American Mathematical Society, 2002.

[17] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2001.

[18] D. Bezdek and K. Bezdek. Shortest billiard trajectories. *Geometriae Dedicata*, 141:197–206, 2009.

[19] G. Birkhoff. Three observations on linear algebra. *Univ. Nac. Tacuman, Rev. Ser. A*, 5:147–151, 1946.

[20] P. V. M. Blagojević, M. Harrison, S. Tabachnikov, and G. M. Ziegler. Counting periodic trajectories of Finsler billiards. *Symmetry, Integrability and Geometry: Methods and Applications (SIGMA)*, 16, 2020.

[21] K. J. Boroczky, E. Makai Jr, M. Meyer, and S. Reisner. On the volume product of planar polar convex bodies – Lower estimates with stability. *Studia Scientiarum Mathematicarum Hungarica*, 50:159–198, 2013.

[22] J. Bourgain and V. D. Milman. New volume ratio properties for convex symmetric bodies in $\mathbb{R}^n$. *Inventiones mathematicae*, 88:319–340, 1987.

[23] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[24] A. Browder. *Mathematical Analysis: An Introduction*. Undergraduate Texts in Mathematics. Springer New York, 2012.

[25] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Math. Program.*, 120:479–495, 2009.

[26] R. E. Burkard, E. Çela, P. M. Pardalos, and L. S. Pitsoulis. *The Quadratic Assignment Problem*. Springer US, Boston, MA, 1998.

[27] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM J. on Optimization*, 9(4):877–900, 1999.

[28] D. Carlson. What are Schur complements, anyway? *Linear Algebra and its Applications*, 74:257–275, 1986.

[29] G. Darboux. Sur le problème de Pfaff. *Bulletin des Sciences Mathématiques et Astronomiques*, 2e série, 6(1):14–36, 1882.

[30] M. A. De Gosson. *Symplectic Geometry and Quantum Mechanics*. Operator Theory: Advances and Applications. Birkhäuser Basel, 2006.

[31] E. De Klerk and F. Vallentin. On the Turing model complexity of interior point methods for semidefinite programming. *SIAM Journal on Optimization*, 26(3):1944–1961, 2016.

[32] P. J. Dickinson and L. Gijben. On the computational complexity of membership problems for the completely positive cone and its dual. *Comput. Optim. Appl.*, 57(2):403–415, 2014.

[33] F. M. Dopico and C. R. Johnson. Complementary bases in symplectic matrices and a proof that their determinant is one. *Linear Algebra and its Applications*, 419(2):772–778, 2006.

[34] M. Dür and G. J. Still. Interior points of the completely positive cone. *Electronic Journal of Linear Algebra*, 17:48–53, 2008.

[35] I. Ekeland. *Convexity Methods in Hamiltonian Mechanics*. Ergebnisse der Mathematik und ihrer Grenzgebiete : a series of modern surveys in mathematics. Folge 3. Springer-Verlag, 1990.

[36] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Elec. Eng. Dept. Stanford University, 2002.

[37] G. Finke, R. E. Burkard, and F. Rendl. Quadratic assignment problems. In Silvano Martello, Gilbert Laporte, Michel Minoux, and Celso Ribeiro, editors, *Surveys in Combinatorial Optimization*, volume 132 of *North-Holland Mathematics Studies*, pages 61–82. North-Holland, 1987.

[38] M. Giaquinta and S. Hildebrandt. *Calculus of Variations I*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2004.

[39] L. Godinho and J. Natário. *An Introduction to Riemannian Geometry: With Applications to Mechanics and Relativity*. Universitext. Springer International Publishing, 2014.

[40] M. Gotay and J. Isenberg. The symplectification of science (symplectic geometry as a basis of physics and mathematics). *Gazette des Mathématiciens*, 54, 1992.

[41] L. J. Gray and D. G. Wilson. Nonnegative factorization of positive semidefinite nonnegative matrices. *Linear Algebra and its Applications*, 31:119–127, 1980.

[42] M. Gromov. Pseudo holomorphic curves in symplectic manifolds. *Inventiones mathematicae*, 82:307–348, 1985.

[43] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[44] V. Guillemin and S. Sternberg. *Symplectic Techniques in Physics*. Cambridge University Press, 1990.

[45] E. Gutkin. Billiard dynamics: An updated survey with the emphasis on open problems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(2):026116, 2012.

[46] E. Gutkin and S. Tabachnikov. Billiards in Finsler and Minkowski geometries. *Journal of Geometry and Physics*, 40(3-4):277–301, 2002.

[47] P. Haim-Kislev. On the symplectic size of convex polytopes. *Geometric and Functional Analysis*, 29(2):440–463, 2019.

[48] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020.

[49] H. Hofer and I. Ekeland. Symplectic topology and Hamiltonian dynamics. *Mathematische Zeitschrift*, 200(3):355–378, 1988/89.

[50] H. Hofer and E. Zehnder. A new capacity for symplectic manifolds. In P. H. Rabinowitz and E. Zehnder, editors, *Analysis, et Cetera*, pages 405–427. Academic Press, 1990.

[51] H. Hofer and E. Zehnder. *Symplectic Invariants and Hamiltonian Dynamics*. Modern Birkhäuser Classics. Springer Basel, 2011.

[52] A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Math. J.*, 20(1):37–39, 1953.

[53] A. Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *American Journal of Mathematics*, 76(3):620–630, 1954.

[54] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2012.

[55] H. Iriyeh and M. Shibata. Symmetric Mahler's conjecture for the volume product in the three dimensional case. *arXiv:1706.01749 [math.MG]*, 2017.

[56] K. Jänich. *Vektoranalysis*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2013.

[57] F. Jarre. On the convergence of the method of analytic centers when applied to convex quadratic programs. *Math. Program.*, 49(1–3):341–358, 1990.

[58] C. R. Johnson. *Matrix Theory and Applications*. AMS Short Course Lecture Notes. American Mathematical Society, 1990.

[59] E. Jones, T. Oliphant, P. Peterson, et al. *SciPy: Open source scientific tools for Python*. `http://www.scipy.org/`, 2001.

[60] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[61] A. B. Katok. Billiard table as a playground for a mathematician. In Victor Prasolov and Yulij Ilyashenko, editors, *Surveys in Modern Mathematics*, London Mathematical Society Lecture Note Series, page 216–242. Cambridge University Press, 2005.

[62] J. Kim and S. Reisner. Local minimality of the volume-product at the simplex. *Mathematika*, 57(1):121–134, 2011.

[63] T. C. Koopmans and M. Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76, 1957.

[64] B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms*. Algorithms and Combinatorics. Springer Berlin Heidelberg, 2007.

[65] S. Krupp and D. Rudolf. Regularity results for shortest billiard trajectories in convex bodies in $\mathbb{R}^n$. *arXiv:2003.00599v3 [math.DS]*, 2020.

[66] S. Krupp and D. Rudolf. Shortest Minkowski billiard trajectories and the computation of the EHZ-capacity of Lagrangian products in $\mathbb{R}^4$. In preparation.

[67] G. Kuperberg. From the Mahler conjecture to Gauss linking integrals. *Geometric and Functional Analysis*, 18:870–892, 2006.

[68] M. Laurent and F. Vallentin. Lecture notes: Semidefinite optimization. Technical report, CWI Amsterdam, Tilburg University and University of Cologne, 2016.

[69] J. M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2003.

[70] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1):193–228, 1998. International Linear Algebra Society (ILAS) Symposium on Fast Algorithms for Control, Signals and Image Processing.

[71] K. Mahler. Ein Minimalproblem für konvexe Polygone. *Mathematica (Zutphen)*, (B7):118–127, 1938.

[72] K. Mahler. Ein Übertragungsprinzip für konvexe Körper. *Časopis pro pěstování matematiky a fysiky*, 068(3-4):93–102, 1939.

[73] J. E. Maxfield and H. Minc. On the matrix equation $X'X = A$. *Proceedings of the Edinburgh Mathematical Society*, 13(2):125–129, 1962.

[74] D. McDuff and D. Salamon. *Introduction to Symplectic Topology*. Oxford University Press, Oxford, UK, 3rd edition, 2017.

[75] S. Mehrotra and J. Sun. A method of analytic centers for quadratically constrained convex quadratic programs. *SIAM J. Numer. Anal.*, 28(2):529–544, 1991.

[76] R. D. C. Monteiro and I. Adler. Interior path following primal-dual algorithms. Part II: Convex quadratic programming. *Math. Program.*, 44(1):43–66, 1989.

[77] M. Nakata. A numerical evaluation of highly accurate multiple-precision arithmetic version of semidefinite programming solver: SDPA-GMP, -QD and -DD. In *2010 IEEE International Symposium on Computer-Aided Control System Design*, pages 29–34, 2010.

[78] M. Nakata, B. J. Braams, K. Fujisawa, M. Fukuda, J. K. Percus, M. Yamashita, and Z. Zhao. Variational calculation of second-order reduced density matrices by strong N-representability conditions and an accurate semidefinite programming solver. *The Journal of Chemical Physics*, 128(16):164113, 2008.

[79] A. Nemirovskii and K. Scheinberg. Extension of Karmarkar's algorithm onto convex quadratically constrained quadratic problems. *Math. Program.*, 72(3):273–289, 1996.

[80] Y. Nesterov and A. Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics, 1994.

[81] Y. Ostrover. When symplectic topology meets Banach space geometry. *Proceedings of the ICM*, 2:959–981, 2014.

[82] C. H. Papadimitriou and K. Steiglitz. Some complexity results for the traveling salesman problem. In *Proceedings of the Eighth Annual ACM Symposium on Theory of Computing*, STOC '76, page 1–9, New York, NY, USA, 1976. Association for Computing Machinery.

[83] M. Radnović. A note on Billiard systems in Finsler plane with elliptic indicatrices. *Publications de l'Institut Mathématique*, 74(88)(94):97–101, 2003.

[84] M. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Mathematical Programming*, 77:129–162, 1997.

[85] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, 2010.

[86] M. G. C. Resende, K. G. Ramakrishnan, and Z. Drezner. Computing lower bounds for the quadratic assignment problem with an interior point algorithm for linear programming. *Operations Research*, 43(5):781–791, 1995.

[87] S. Sahni and T. Gonzalez. P-complete approximation problems. *J. ACM*, 23(3):555–565, 1976.

[88] L. A. Santaló. Un invariante afin para los cuerpos convexos del espacio de n dimensiones. *Portugaliae mathematica*, 8(4):155–161, 1949.

[89] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley and Sons, Inc., USA, 1986.

[90] I. Schur. Über eine Klasse von Mittelbildungen mit Anwendungen auf der Determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22:9–20, 1923.

[91] J. Schur. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *Journal für die reine und angewandte Mathematik*, 140:1–28, 1911.

[92] K. F. Siburg. Symplectic capacities in two dimensions. *Manuscripta mathematica*, 78(2):149–164, 1993.

[93] S. Tabachnikov and Pennsylvania State University. Mathematics Advanced Study Semesters. *Geometry and Billiards*. Student mathematical library. American Mathematical Society, 2005.

[94] R. J. Tykodi. On Euler's theorem for homogeneous functions and proofs thereof. *Journal of Chemical Education*, 59(7):557, 1982.

[95] M. Usher. *Low-dimensional and Symplectic Topology*. Proceedings of symposia in pure mathematics. American Mathematical Society, 2011.

[96] G. Van Rossum and F. L. Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[97] L. Vandenberghe. *The CVXOPT linear and quadratic cone program solvers*. `https://www.seas.ucla.edu/~vandenbe/publications/coneprog.pdf`, 2010.

[98] C. Viterbo. Capacités symplectiques et applications. In *Séminaire Bourbaki, no.714, volume 1988/89, exposés 700-714*, number 177/178 in Astérisque, pages 345–362. Société mathématique de France, 1989.

[99] C. Viterbo. Metric and isoperimetric problems in symplectic geometry. *Journal of the American Mathematical Society*, 13(2):411–431, 2000.

[100] A. Weinstein. Symplectic geometry. *Bull. Amer. Math. Soc. (N.S.)*, 5(1), 1981.

[101] Y. Yajima. Convex envelopes in optimization problems. In *Encyclopedia of Optimization*, pages 343–344. Springer US, Boston, MA, 2001.

[102] M. Yamashita, K. Fujisawa, K. Nakata, M. Nakata, M. Fukuda, K. Kobayashi, and K. Goto. *A high-performance software package for semidefinite programs: SDPA 7*. Research Report B-463, Dept. of Mathematical and Computing Science, Tokyo Institute of Technology, Tokyo, Japan, 2010.

[103] P. F. Zhang, F. Zhang, S. J. Axler, and F. W. Gehring. *Matrix Theory: Basic Results and Techniques*. Universitext (Berlin. Print). Springer, 1999.

[104] Q. Zhao, S. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *Journal of Combinatorial Optimization*, 2, 1997.

[105] G. M. Ziegler. *Lectures on Polytopes*. Graduate Texts in Mathematics. Springer New York, 1995.

Publication list

In preparation

- Stefan Krupp, Daniel Rudolf. *Shortest Minkowski billiard trajectories and the computation of the EHZ-capacity of Lagrangian products in* $\mathbb{R}^4$.

Published

- Stefan Krupp, Daniel Rudolf. *Regularity results for shortest billiard trajectories in convex bodies in* $\mathbb{R}^n$. *arXiv:2003.00599v3 [math.DS]*, 2020.

Surveys

- Maria Dostert, Stefan Krupp, Jan Rolfes. *Das Problem der Kugelpackung.* Snapshots of modern mathematics from Oberwolfach, 2016, https://imaginary.org/snapshot/das-problem-der-kugelpackung

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Frank Vallentin betreut worden.

_____

Stefan Krupp

Teilpublikationen:

- Stefan Krupp, Daniel Rudolf. *Regularity results for shortest billiard trajectories in convex bodies in* $\mathbb{R}^n$. *arXiv:2003.00599v3 [math.DS]*, 2020.
- Stefan Krupp, Daniel Rudolf. *Shortest Minkowski billiard trajectories and the computation of the EHZ-capacity of Lagrangian products in* $\mathbb{R}^4$. In preparation.

Lebenslauf

Persönliche Daten:

Name	Stefan Krupp
Geboren	10.06.1992 in Linz/Rhein
Staatsangehörigkeit	deutsch

Studienverlauf:

Seit 10/2017	Promotionsstudium an der Universität zu Köln Betreuer: Prof. Dr. Frank Vallentin
10/2014 – 09/2017	Studium an der Universität zu Köln Fachrichtung: Wirtschaftsmathematik Abschluss: Master of Science
10/2011 – 09/2014	Studium an der Universität zu Köln Fachrichtung: Wirtschaftsmathematik Abschluss: Bachelor of Science

Schulbildung:

08/2008 – 06/2011	Ludwig-Erhard-Berufskolleg Bonn Abschluss: Abitur
08/2002 – 06/2008	Realschule Linz

Beruflicher Werdegang:

Seit 04/2017	Wissenschaftlicher Mitarbeiter Universität zu Köln
05/2016 – 09/2016	Wissenschaftliche Hilfskraft Universität zu Köln
03/2015 – 03/2016	Wissenschaftliche Hilfskraft Universität zu Köln