National Research University Higher School of Economics

Faculty of Mathematics

as a manuscript

Ivanova Anastasiia

Geometric approach to interior-point methods

Summary of the PhD thesis
for the purpose of obtaining academic degree
Doctor of Philosophy in Mathematics

Academic supervisors:
Roland Hildebrand,
Doctor of Sciences
and Vladimir Protasov,
Doctor of Sciences

Moscow - 2024

1 Introduction

Interior point methods were originally developed as an alternative to the simplex method for solving linear programming problems. Unlike the latter, they generate a sequence of points for which all inequality constraints are strictly fulfilled. Actually, methods generating a sequence of interior points of a feasible set have been used for a long time to solve nonlinear problems. By interior point methods we will understand here the methods that explicitly or implicitly use a barrier on the feasible set or on the cone underlying the problem. In the case of linear programming, this cone is a non-negative orthant.

The first to propose such a method was [Dikin, 1967]. In principle, this method defined the next point as an argument of the minimum of the linear price function on Dikin's ellipsoid with the center in the previous iteration. However, this method had no guarantees on the convergence rate, and it required a strictly feasible initial point. Later [Karmarkar, 1984] constructed the first interior point method for linear programming, which was proved to have polynomial complexity. This method used a projective transformation of the feasible set at each step before constructing an ellipsoid inscribed in that set. Since the class of linear functions is not invariant with respect to projective transformations, one had to extend this class to fractional linear functions, which were minimized on the constructed ellipsoids. This shortcoming of Karmarkar's method motivated the development of variants of methods using affine transformations, which eventually led to the rediscovery of Dikin's method ([Barnes, 1986], [Vanderbei et al., 1986]). All later versions of Karmarkar's method could compete with the simplex method in practice, but had no theoretical guarantees of convergence rate. Note that Karmarkar himself has already pointed out that, on the one hand, his method can be interpreted as a barrier method, i.e. minimizing the weighted sum of the initial cost function and the standard logarithmic barrier on the orthant, and, on the other hand, as a potential reduction method.

In a series of papers, D. Bayer and J. Lagarias studied in detail the relationship between Karmarkar-type methods and barrier methods. In particular, the methods were classified into projective- and affine-scaling methods, depending on which transformations are applied to the feasible set. In the future, mainly affine scaling methods were developed. The notion of central path was also introduced. [Renegar, 1988] constructed a path-following method which produced a sequence of points close to the central path and which was provably of polynomial complexity. Primal-dual methods were proposed in [Kojima et al., 1989] and [Monteiro and Adler, 1989]. In [Kojima et al., 1993] and [Mizuno et al., 1995] infeasible primal-dual methods were proposed, with iterations for which equality-type constraints might not be satisfied. Potential reduction methods for linear programming were proposed in [Tanabe, 1988] and [Todd and Ye, 1990]. The Tanabe-Todd-Ye potential proposed in these works has formed the basis of more recent, advanced potential reduction methods for semi-definite programming.

In the late 1980s, Yu. Nesterov and A. Nemirovsky generalized the interior point methods for linear programming to optimization problems with arbitrary convex conic constraints and introduced the concepts of a conic program and a self-concordant barrier. In particular, they proposed polynomial complexity methods for solving semi-definite programs. In [Nesterov and Nemirovsky, 1992] they also constructed the theory of conic duality, which generalizes duality in linear programs. The complete theory is described in the monograph [Nesterov and Nemirovskii, 1994], which describes both central path following and potential-reducing methods. Later, [Nesterov, 1997] found that the methods of potential reduction take steps similar to the long steps in the path-following methods.

The methods for solving linear programs were independently generalised to the case of semi-definite programming in a series of works by [Alizadeh, 1995]. However, his methods used scaling automorphisms of a cone and in principle could not be used for solving problems over non-symmetric cones.

The next phase was the development of the theory of autoscaled barriers and the long-step methods based on this theory, which exploited the rich structure of symmetric cones ([Monteiro, 1997, Monteiro and Zhang, 1998, Nesterov and Todd, 1997, Nesterov and Todd, 1998, Tunçel, 2000]). This class of methods has proved to be the most effective in practice and is still used in solvers today [Todd et al., 1998] and [Toh et al., 1999].

Note that the special role of symmetric cones as the most natural generalization of the non-negative orthant in conic programming was first noted by L. Faibuzovich, who explicitly used their algebraic structure to construct algorithms ([Faybusovich, 1997a] and [Faybusovich, 1997b]). While the property of autoscalability is inherent to a barrier, symmetry is a property of a cone. In the early 2000s it was discovered that these two concepts are closely related and that autoscaled barriers exist on all symmetric cones and only on them. An overview of these developments is presented in [Hauser and Güler, 2002].

In [Güler, 1997], methods based on the autoscaling property were generalized to the class of hyperbolic cones, and in [Chua, 2009] to the class of homogeneous cones. In [Nesterov, 2012], the notion of a scaling point was also defined for arbitrary self-concordant barriers.

In the last decade, several new universal constructions of self-concordant barriers have been published for cones and convex sets ([Bubeck and Eldan, 2019, Hildebrand, 2014, Lee and Yue, 2021]), as well as for linear programming ([Lee, 2016][Section 6.3]). In [Abernethy and Hazan, 2016], a relationship between path-following methods using the entropic barrier and the simulated annealing algorithm was found. In [Nesterov and Tunçel, 2016], instead of the auto-scalability property, the weaker negative curvature property, which is fulfilled in particular for standard barriers on hyperbolic cones, was used to construct algorithms. In recent years, there has been a growing interest in barriers on asymmetric cones, e.g., the exponential cone ([Dahl and Andersen, 2022]).

Note that along with the interior point methods, generalizations of the simplex method to semidefinite programming have been developed, in particular, in the series of works by V. Zhadan.

The results of this thesis are published as the following two articles:

1. Ivanova, A. and Hildebrand, R. (2023). Optimal step length for the maximal decrease of a self-concordant function by the newton method. *Optimization Letters*, pages 1–8

2. Hildebrand, R. and Ivanova, A. (2022). Extremal cubics on the circle and the 2-sphere. *Results in Mathematics*, 77(3):1–33

We outline the rest of the manuscript and highlight the main contributions.

- Chapter 2, which is described in Section 2 of this summary, is devoted to applying optimal control theory to find the optimal step size for Newton's method and the path-following method on the class of self-concordant functions. For the Newton method our goal is to find an optimal step length in terms of the functional value, i.e. this optimal step-size should maximize the decrease of the functional value. For the path-following method the quality of the current point x_0 is measured by the distance in the local metric to the straight line approximating the central path. Our goal is to move as far along the straight line as possible without losing in quality in terms of the distance to the central path. This leads to a problem: how large this distance can be if we pass from x_0 to another point x_1. In both problems we consider the evolution of values (function and its derivatives) over the whole segment between starting and endpoints. Thus, we are looking at an infinite-dimensional object and a suitable apparatus for solving this problem is optimal control theory. So, we state these problems as optimal control problems and use optimal control theory to solve them.

- In Chapter 3, which is described in Section 3 of this summary, we consider homogeneous cubic polynomials on the unit sphere. We consider the set of cubics which are bounded by 1 on the unit sphere as a 10-dimensional convex body, since each cubic has 10 coefficients. Our main goal is to provide a complete classification of the extremal points of this body. Since homogeneous cubic polynomials in 3 variables have 10 independent coefficients, we can define all these coefficients if we fix 4 maxima of this polynomial on a sphere. Each maximum point corresponds to 3 optimality conditions, therefore 4 maxima define a system of 12 linear equations. Note that P is invariant with respect to rotation in $\mathbb{R}^3$. That is, if we rotate the sphere, then the 4 points on the sphere rotate too, and this new point also corresponds to some extreme polynomial. It turns out that instead of values of maxima vectors, we should consider angles between these vectors, since they are invariant with respect to rotation. Solving this problem, we get a representation of the cubic polynomialas a function of the Gramian of the maxima points, i.e. in the general case an extreme polynomial is more elegantly described by the maxima, or their Gramian, than by the coefficients themselves. Moreover, we consider some special cases and provide a full classification of the extremal cubics. The motivation for this investigation is that the unit norm ball defined by condition $\sum_{r,s,t=1}^{n} P_{rst}x_r x_t x_s \leq 1$ for all $x \in \mathbb{R}^n, ||x|| = 1$ in the space of polynomials is used as set of admissible controls in the control problem formulated for finding the optimal step in the path-following methods, and its extreme points define the possible optimal controls.

- In Chapter 4, which is described in Section 4 of this summary, we discuss some future directions of this work.

2 Optimal control approach to find the optimal step length

This chapter is devoted to the optimal control approach to find the optimal step for the Newton method and for the path-following method. The idea to use optimal control methods for the worst-case analysis of first order methods has been developed by Laurent Lessard and co-authors in [Lessard et al., 2016], where they use robust control techniques related to the search for quadratic Lyapunov functions to derive numerical upper bounds on convergence rates for the Gradient method, the Heavy-ball method, Nesterov's accelerated method, and related variants by solving small, simple semi-definite programming problems. The same technique was also considered in [Taylor et al., 2018, Hu and Lessard, 2017]. In contrast to this finite-dimensional approach we shall use optimal control theory in the infinite-dimensional setting of the Pontryagin maximum principle.

The first part of the chapter is devoted to the problem of finding the optimal step length of Newton's method on the class of self-concordant functions, motivated by the appearance of this class in barrier methods for conic programming, in particular, when solving linear programs, second-order cone programs, and semi-definite programs.

Definition 2.1. A convex C^3 function $f : D \to \mathbb{R}$ on a convex domain D is called self-concordant if it satisfies the inequality

$$|f'''(x)[h,h,h]| \leq 2(f''(x)[h,h])^{3/2} \tag{1}$$

for all $x \in D$ and all tangent vectors h.

Step lengths for the damped Newton method were also considered in [Burdakov, 1980, Ralph, 1994, Nesterov, 2018]. The behaviour of the Newton decrement and the function value under specific step sizes has been studied in [Renegar, 2001, Section 2.2]. In [De Klerk et al., 2020a, Corollary 6.1] the decrease of the distance from the optimum and of the norm of the gradient, both in the local metric of the initial point, have been bounded for self-concordant functions if the initial point is close enough to the minimum. The same bound has been obtained in [De Klerk et al., 2020a, Corollary 6.3] for the local metric of the minimum. A bound on the decrease in function value can be derived from [De Klerk et al., 2020a, Theorem 5.3], however, it depends on the difference between the current and optimal function values. In that paper an inexact Newton step can be taken, and the bound depends on the error. The methods used in [De Klerk et al., 2020a] rely on semi-definite programming (see also [Drori and Teboulle, 2014]) and are completely different from those employed here.

In this chapter we find the optimal step length of Newton's method with respect to the decrease of the function value. This criterion was considered in [Nesterov and Nemirovskii, 1994, Theorem 2.2.1], where the decrease has been lower bounded by an explicit function of the step length γ_k and the Newton decrement ρ_k. The same bound has been derived in [Gao and Goldfarb, 2019] in a more general context. In the latter paper it is shown that the step length $\gamma_k = \frac{1}{1+\rho_k}$ maximizes this lower bound. The same expression for the step length is also proposed in [Nesterov and Nemirovskii, 1994, Theorem 2.2.3] for larger values of the decrement. While in [Gao and Goldfarb, 2019], and implicitly in [Nesterov and Nemirovskii, 1994], the step length has been obtained as the maximizer of a bound, in the present paper we show by employing optimal control theory that this step length is actually optimizing the function value itself. It turns out, however, that no further improvement over the results in the mentioned papers occurs, despite the use of the exact criterion.

Optimal control theory has already been used in [Hildebrand, 2021] to find an optimal step-length γ^* for the Newton method on self-concordant functions. However, a different strategy has been adopted there. Instead of the worst-case function value, as in the present work, the worst-case Newton decrement in the next iteration is minimized. This criterion is more in line with the philosophy of interior-point methods as presented in [Nesterov and Nemirovskii, 1994], but it has the drawback that if the decrement is larger than 1, no progress can be guaranteed at all. Also, the optimal value of the step length turns out to be not expressible in closed form in general. The criterion used in the present paper, on the contrary, can be strictly improved at each step, no matter how far we are from the optimum at the current iteration, and the value of the optimal step length is a simple analytic function of the data available at the current iteration.

The second part of the chapter is devoted to the problem of finding the optimal step for the path-following method on the class of self-concordant functions. The distance to the central path was also considered in [Nesterov and Todd, 1998]. The distance to the central path in the local metric was chosen as the criterion.

2.1 Optimal step length for the maximal decrease of a self-concordant function by the Newton method

In this chapter we consider Newton's method with a damped step, producing iterations according to

$$x_{k+1} = x_k - \gamma_k (F''(x_k))^{-1} F'(x_k), \tag{2}$$

where $\gamma_k \in (0,1]$ is the step-size and $\gamma_k = 1$ corresponds to a full step.

The authors of [Nesterov and Nemirovskii, 1994] describe the state at iteration k by a single scalar, the *Newton decrement*

$$\rho_k = ||F'(x_k)||_{F''(x_k)} := \sqrt{(F'(x_k))^\top (F''(x_k))^{-1} F'(x_k)}. \tag{3}$$

In this section consider the problem of finding a step length γ_k which maximizes the decrease $F(x_k) - F(x_{k+1})$ of the function value in the worst case realization of the function $F(\cdot)$. So, we firstly need for given step length and given decrement to find the worst realization of the function giving the minimal decrease, and then to maximize this progress over the value of the step length, yielding the optimal step length as a function of the decrement. This leads to the following optimization problem:

$$\max_{\gamma_k} \min_{F \in \mathcal{S}} \left(F(x_k) - F(x_{k+1}) \right), \tag{4}$$

where γ_k is the step length, x_{k+1} is given by (2), the decrement $||F'(x_k)||_{F''(x_k)}$ is fixed to some value ρ_k, and $\mathcal{S}$ is the class of functions satisfying (1).

To solve problem (4) we consider a single iteration of the Newton method. Let the end point x_{k+1} be given by (2) and consider the line segment between x_k and x_{k+1}. We study the evolution of the values of the function and its derivatives along this segment. This distinguishes our approach from the approach in [De Klerk et al., 2020a], where n iterations and only the values of the function and its derivatives at the points $x_1, \dots, x_n$, i.e., a finite dimensional object, are considered. In contrast to this we consider an infinite dimensional object. The suitable apparatus to solve this problem is optimal control theory. We can solve this problem analytically. As a result we get that $\gamma^\star = \frac{1}{1+\rho}$ is the optimal step. The same step length was already proposed in [Nesterov and Nemirovskii, 1994], and in [Gao and Goldfarb, 2019] it is shown that this step-size maximizes a lower bound on the decrease of the function value. In our work we have proved that this step length is actually optimal for this criterion.

2.2 Optimal step for the path following method

This chapter is devoted to the problem of conditional minimisation of a linear functional on a given feasible set. Assume that we need to solve the following optimization problem

$$\min_{x \in X} \langle c, x \rangle,$$

where X is a convex closed set with non-empty interior D not containing a straight line, equipped with a self-concordant barrier $F : D \to \mathbb{R}$. Here $c \in \mathbb{R}_n = (\mathbb{R}^n)^*$ is a non-zero linear functional on D. We assume the existence of a solution x^* of this problem. As we described in Section **??**, instead of the initial problem we can consider the following family of problems, parameterized by a real parameter $\tau \in \mathbb{R}_+$:

$$\min_x \tau\langle c, x\rangle + F(x).$$

Thus we pass to the problem of unconditional minimization where the objective functional is strongly self-concordant. Therefore, to solve these auxiliary problems we can use Newton's method.

For large enough τ the minimizers $x^*(\tau)$ of the auxiliary problem exist and are unique. Moreover, they are differentiable with respect to the parameter τ and form a curve called *the central path.* The approximation of the central path at the point $\hat{x}$ is given by the affine line

$$x^*(\tau, \hat{x}) = \hat{x} - (F''(\hat{x}))^{-1}(F'(\hat{x}) + \tau c).$$

The approximation lies on a straight line with direction

$$p(\hat{x}) = -(F''(\hat{x}))^{-1}c$$

and passes through the point $\hat{x} + r(\hat{x})$, where

$$r(\hat{x}) = -(F''(\hat{x}))^{-1}(F'(\hat{x}) + \tau_0 c)$$

and τ_0 is arbitrary.

The quality of the current point $\hat{x}$ is measured by the distance in the local metric to the straight line approximating the central path. So, the nearest point to $\hat{x}$ on this line, measured in the local metric, has parameter

$$\begin{aligned}\tau^*_{\hat{x}} &= \arg\min_{\tau\in\mathbb{R}} \|x^*(\tau,\hat{x}) - \hat{x}\|_{\hat{x}} = \arg\min_{\tau\in\mathbb{R}} \|(F''(\hat{x}))^{-1}\left(F'(\hat{x}) + \tau c\right)\|_{\hat{x}} \\ &= \arg\min_{\tau\in\mathbb{R}} \sqrt{(F'(\hat{x}) + \tau c)\,(F''(\hat{x}))^{-1}\,(F'(\hat{x}) + \tau c)} = \frac{c^T(F''(\hat{x}))^{-1}F'(\hat{x})}{c^T(F''(\hat{x}))^{-1}c}.\end{aligned}$$

The corresponding distance to the approximation of the central path is given by

$$\begin{aligned}\delta(\hat{x}) &= \|(F''(\hat{x}))^{-1}\left(F'(\hat{x}) + \tau^*_{\hat{x}}c\right)\|_{\hat{x}} \\ &= \sqrt{(F'(\hat{x}))^T(F''(\hat{x}))^{-1}F'(\hat{x}) - \frac{c^T(F''(\hat{x}))^{-1}F'(\hat{x})}{c^T(F''(\hat{x}))^{-1}c}}.\end{aligned}$$

Note that the general criterion is to measure the distance to a target point on the central path, but in our work we measure the distance to the central path, that is, we do not fix the target point.

Our goal is to move as far along the straight line as possible without losing quality in terms of the distance to the central path. This leads to a problem: **how large this distance can be if we pass from x_0 to another point x_1.** An approximation of the central path for the different points is represented in Figure 1. In this figure, the black point corresponds to the starting point, and the black line corresponds to the approximation of the central path at that point. We start moving from the current point, but the line itself moves with the point. For example, if we step to the blue point, which is close enough to the initial approximation of the central path, this new point has its own approximation of the central path, the blue line. The distance to this line from the blue point is larger. But it is important for us that this distance is not worse than the distance from the starting point to the initial approximation. If we step too far then the approximation will change a lot and the distance can increase a lot. If we step too close, we will move slower than we could, that is, it will be slower to converge to the solution. So the goal is to figure out how far we can move along the central path without losing in quality at the new point.

In [De Klerk et al., 2020b] authors also study the quality of iterations of the path-following method, but they consider the current and the next point and the values of function and its derivatives at these points. Moreover, they consider several iterations, whereas we consider only one iteration. We consider not only values at the endpoints but the evolution of values (function and its derivatives) over the whole segment between starting and endpoints. Thus, we are looking at an infinite-dimensional object, whereas in [De Klerk et al., 2020b] a finite-dimensional object has investigated. Therefore, we

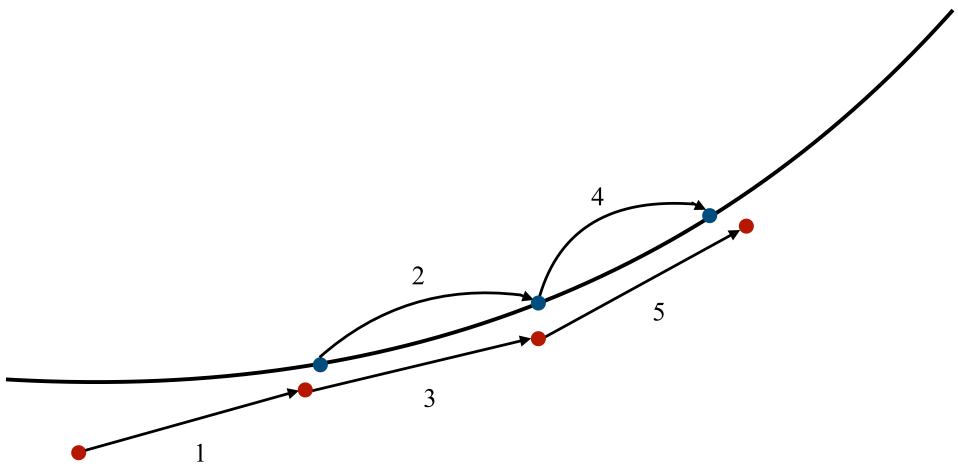

Figure 1: An approximation of the central path for the different points.

use more information that can help us to estimate an optimal step length more precisely. To this end, we shall investigate the evolution of the derivatives of f along the segment between x_0 and x_1, which leads to an infinite-dimensional problem and a suitable apparatus for solving this problem is optimal control theory. We formulate this problem as an optimal control problem and further numerically solve it.

3 Extremal cubics on the circle and the 2-sphere

The problem of maximizing a homogeneous cubic over the unit sphere is NP-hard and arises in various applications in non-convex and combinatorial optimization, e.g., the stable set problem which was considered in [Nesterov et al., 2003]. Several approaches for the solution of this problem have been proposed in [So, 2011, Nie, 2012, Zhang et al., 2012, Ahmed and Still, 2019, Buchheim et al., 2019, de Klerk and Laurent, 2020, Fang and Fawzi, 2021].
The problem is equivalent to determining the intersection of a given ray in the space of homogeneous cubics with the boundary of the unit norm ball in this space

$$B_1(S^{n-1}) := \{p \in \mathbb{R}[x_1, \dots, x_n]_3 \mid \|p\| \leq 1\},$$

where the norm is defined by

$$\|p\| = \max_{\|x\|_2=1} p(x), \tag{5}$$

and $\|\cdot\|_2$ is the usual Euclidean norm in $\mathbb{R}^n$. A semi-definite description of this norm ball for $n = 3$ follows from Propositions 4.8 and 5.1 in [Saunderson, 2019], for $n = 2$ such a description is readily constructed from the semi-definite description of the cone of nonnegative univariate trigonometric polynomials [Nesterov, 2000, Section 3.4]. Inequivalent semi-definite descriptions which can be generalized to balls of non-homogeneous cubics can be found in [Hildebrand, 2022].

In this work we analyze the corresponding semi-definite representable norm balls in the space of homogeneous cubics, which have dimensions 10 and 4, respectively. In particular, we determine all their extreme points and study their facial structure.

Extremal nonnegative polynomials have among others been studied in [Choi and Lam, 1977, Reznick, 1978, Reznick, 2000, Naldi, 2014]. Extremal homogeneous cubics in three variables which are nonnegative on the orthant have been studied in [Ando, 2021]. The facial structure of cones of nonnegative polynomials has been studied in [Kunert, 2014, Blekherman et al., 2015]. In particular, the exposed faces of the cone of nonnegative polynomials have been characterized as the set of nonnegative polynomials which vanish on a certain subset of points [Blekherman et al., 2013, Section 4.4.5], [Kunert, 2014, Prop. 1.69].

Our motivation for studying the structure of the norms balls $B_1(S^{n-1})$ for $n = 2, 3$ is that in the analysis of optimization algorithms on self-concordant functions by methods of optimal control these balls appear as control sets, and the optimal controls are extreme points of these balls [Hildebrand, 2021].

In this work we define some basic notations, in particular, faces of a convex set, and investigate the connection between the faces of the norm ball $B_1(S^{n-1})$ and the maxima of cubics on S^{n-1}, after we consider the norm balls $B_1(S^1)$ and $B_1(S^2)$. For $n = 2$ we completely describe the facial structure of this norm ball, while for $n = 3$ we classify all extremal points and describe some families of faces.

4 Conclusion

Interior point methods have a history of more than 35 years and their appearance made a major breakthrough in convex optimization. These methods are still actively used for solving large-scale optimization problems and remain out of competition.

One of the important practical aspects of implementing any algorithm is the appropriate choice of step length. This problem for interior point methods is the focus of this thesis. In this work, we propose to use an approach in which the problem of finding the optimal step of the method is formulated as an optimal control problem and then optimal control theory is applied to solve it. In Chapter **??**, we gave the history of interior point methods and summarised the main theoretical concepts that were used in the study, among them Newton's method, self-concordant functions, path-following methods and optimal control theory.

In Chapter 2, we considered the problem of finding the optimal step or iterate for optimisation methods using optimal control theory. In the first part of the chapter we consider the problem of finding the optimal step length for the Newton method on the class of self-concordant functions, with the decrease in function value as criterion. The second part is devoted to finding an optimal step of the path-following method when minimizing a linear function using a self-concordant barrier. The quality of the current point x_0 is measured by the distance in the local metric to the straight line approximating the central path. Our goal is to move as far along the straight line as possible without losing in quality in terms of the distance to the central path. This leads to a problem: how large this distance can be if we pass from x_0 to another point x_1. We formulate both problems as optimal control problems and use optimal control theory to solve them.

In Chapter 3, we study balls of homogeneous cubics on $\mathbb{R}^n$, $n = 2, 3$, which are bounded by unity on the unit sphere. For $n = 2$ we completely describe the facial structure of this norm ball, while for $n = 3$ we classify all extremal points and describe some families of faces. The motivation for studying this problem is that these balls of polynomials appear as control sets in the aforementioned optimal control problems, and their extreme points are candidates for the optimal control.

4.1 Future directions

In this thesis we solved the problem of finding the optimal iterate for a path-following method numerically in two dimensions. If the space where the initial optimization problem takes place has dimension $n \geq 3$, then the optimal control problem arising for the corresponding task of finding the optimal iterate has a similar structure, but the variables y, q are 3-dimensional vectors. The general dimension n hence reduces to dimension 3.

One of the main directions of development of this work is hence to solve the problem of finding the optimal step of the central path method using self-concordant barriers in the three-dimensional case. Contrary to the two-dimensional case, in the three-dimensional case the optimal control set is a set of homogeneous cubic polynomials bounded by one on the sphere. As a first step towards solving this problem, we solved the problem of classifying the extremal points of this set; the complete classification is presented in Chapter 3. The optimal control problem itself, which can be solved also only numerically, could not be considered in the framework of this thesis and is subject to future research.

References

[Abernethy and Hazan, 2016] Abernethy, J. and Hazan, E. (2016). Faster convex optimization: Simulated annealing with an efficient universal barrier. In *International Conference on Machine Learning*, pages 2520–2528. PMLR.

[Ahmed and Still, 2019] Ahmed, F. and Still, G. (2019). Maximization of homogeneous polynomials over the simplex and the sphere: Structure, stability, and generic behavior. *J. Optimiz. Theory App.*, 181:972–996.

[Alizadeh, 1995] Alizadeh, F. (1995). Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM journal on Optimization*, 5(1):13–51.

[Ando, 2021] Ando, T. (2021). On extremal positive semidefinite forms of cubic homogeneous polynomials of three variables. *arXiv preprint arXiv:2109.01319*.

[Barnes, 1986] Barnes, E. R. (1986). A variation on karmarkar's algorithm for solving linear programming problems. *Mathematical programming*, 36:174–182.

[Blekherman et al., 2015] Blekherman, G., Iliman, S., and Kubitzke, M. (2015). Dimensional differences between faces of the cones of nonnegative polynomials and sums of squares. *Int. Math. Res. Not.*, 2015:8437–8470.

[Blekherman et al., 2013] Blekherman, G., Parrilo, P. A., and Thomas, R. R., editors (2013). *Semidefinite Optimization and Convex Algebraic Geometry.* MOS-SIAM series on Optimization. SIAM.

[Bubeck and Eldan, 2019] Bubeck, S. and Eldan, R. (2019). The entropic barrier: Exponential families, log-concave geometry, and self-concordance. *Mathematics of Operations Research*, 44(1):264–276.

[Buchheim et al., 2019] Buchheim, C., Fampa, M., and Sarmiento, O. (2019). Tractable relaxations for the cubic one-spherical optimization problem. In *Optimization of Complex Systems: Theory, Models, Algorithms and Applications*, volume 991 of *Advances in Intelligent Systems and Computing*, pages 267–276. Springer.

[Burdakov, 1980] Burdakov, O. P. (1980). Some globally convergent modifications of Newton's method for solving systems of nonlinear equations. *Doklady Akademii Nauk*, 254(3):521–523.

[Choi and Lam, 1977] Choi, M.-D. and Lam, T.-Y. (1977). Extremal positive semidefinite forms. *Math. Ann.*, 231:1–18.

[Chua, 2009] Chua, C. B. (2009). At-algebraic approach to primal-dual interior-point algorithms. *SIAM Journal on Optimization*, 20(1):503–523.

[Dahl and Andersen, 2022] Dahl, J. and Andersen, E. D. (2022). A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization. *Mathematical Programming*, 194(1):341–370.

[De Klerk et al., 2020a] De Klerk, E., Glineur, F., and Taylor, A. B. (2020a). Worst-case convergence analysis of inexact gradient and Newton methods through semidefinite programming performance estimation. *SIAM Journal on Optimization*, 30(3):2053–2082.

[De Klerk et al., 2020b] De Klerk, E., Glineur, F., and Taylor, A. B. (2020b). Worst-case convergence analysis of inexact gradient and newton methods through semidefinite programming performance estimation. *SIAM Journal on Optimization*, 30(3):2053–2082.

[de Klerk and Laurent, 2020] de Klerk, E. and Laurent, M. (2020). Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere. Published online in Math. Program.

[Dikin, 1967] Dikin, I. (1967). Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk*, 174(4):747–748.

[Drori and Teboulle, 2014] Drori, Y. and Teboulle, M. (2014). Performance of first-order methods for smooth convex minimization: a novel approach. *Math. Program.*, 145(1–2):451–482.

[Fang and Fawzi, 2021] Fang, K. and Fawzi, H. (2021). The sum-of-squares hierarchy on the sphere and applications in quantum information theory. *Math. Program.*, 190:331–360.

[Faybusovich, 1997a] Faybusovich, L. (1997a). Euclidean jordan algebras and interior-point algorithms. *Positivity*, 1:331–357.

[Faybusovich, 1997b] Faybusovich, L. (1997b). Linear systems in jordan algebras and primal-dual interior-point algorithms. *Journal of computational and applied mathematics*, 86(1):149–175.

[Gao and Goldfarb, 2019] Gao, W. and Goldfarb, D. (2019). Quasi-Newton methods: superlinear convergence without line searches for self-concordant functions. *Optimization Methods and Software*, 34(1):194–217.

[Güler, 1997] Güler, O. (1997). Hyperbolic polynomials and interior point methods for convex programming. *Mathematics of Operations Research*, 22(2):350–377.

[Hauser and Güler, 2002] Hauser, R. A. and Güler, O. (2002). Self-scaled barrier functions on symmetric cones and their classification. *Foundations of Computational Mathematics*, 2:121–143.

[Hildebrand, 2014] Hildebrand, R. (2014). Canonical barriers on convex cones. *Mathematics of operations research*, 39(3):841–850.

[Hildebrand, 2021] Hildebrand, R. (2021). Optimal step length for the Newton method: Case of self-concordant functions. *Math. Methods. Oper. Res.*, 94:253–279.

[Hildebrand, 2022] Hildebrand, R. (2022). Semi-definite representations for sets of cubics on the 2-sphere. *J. Optim. Theory Appl.*, 195:666–675.

[Hu and Lessard, 2017] Hu, B. and Lessard, L. (2017). Control interpretations for first-order optimization methods. In *2017 American Control Conference (ACC)*, pages 3114–3119. IEEE.

[Karmarkar, 1984] Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311.

[Kojima et al., 1993] Kojima, M., Megiddo, N., and Mizuno, S. (1993). A primal—dual infeasible-interior-point algorithm for linear programming. *Mathematical programming*, 61(1):263–280.

[Kojima et al., 1989] Kojima, M., Mizuno, S., and Yoshise, A. (1989). *A primal-dual interior point algorithm for linear programming*. Springer.

[Kunert, 2014] Kunert, A. (2014). *Facial structure of cones of nonnegative forms*. PhD thesis, University Konstanz, Konstanz.

[Lee, 2016] Lee, Y. T. (2016). *Faster algorithms for convex and combinatorial optimization*. PhD thesis, Massachusetts Institute of Technology.

[Lee and Yue, 2021] Lee, Y. T. and Yue, M.-C. (2021). Universal barrier is n-self-concordant. *Mathematics of Operations Research*, 46(3):1129–1148.

[Lessard et al., 2016] Lessard, L., Recht, B., and Packard, A. (2016). Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1):57–95.

[Mizuno et al., 1995] Mizuno, S., Kojima, M., and Todd, M. J. (1995). Infeasible-interior-point primal-dual potential-reduction algorithms for linear programming. *SIAM Journal on Optimization*, 5(1):52–67.

[Monteiro, 1997] Monteiro, R. D. (1997). Primal–dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7(3):663–678.

[Monteiro and Adler, 1989] Monteiro, R. D. and Adler, I. (1989). Interior path following primal-dual algorithms. part i: Linear programming. *Mathematical programming*, 44(1):27–41.

[Monteiro and Zhang, 1998] Monteiro, R. D. and Zhang, Y. (1998). A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Mathematical Programming*, 81(3):281–299.

[Naldi, 2014] Naldi, S. (2014). Nonnegative polynomials and their Carathéodory number. *Discrete & Computational Geometry*, 51:559–568.

[Nesterov, 1997] Nesterov, Y. (1997). Long-step strategies in interior-point primal-dual methods. *Mathematical Programming*, 76(1):47–94.

[Nesterov, 2000] Nesterov, Y. (2000). Squared functional systems and optimization problems. In Frenk, H., Roos, K., Terlaky, T., and Zhang, S., editors, *High Performance Optimization*, chapter 17, pages 405–440. Kluwer Academic Press, Dordrecht.

[Nesterov, 2012] Nesterov, Y. (2012). Towards non-symmetric conic optimization. *Optimization methods and software*, 27(4-5):893–917.

[Nesterov, 2018] Nesterov, Y. (2018). *Lectures on Convex Optimization*, volume 137 of *Springer Optimization and its Applications*. Springer.

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

[Nesterov and Nemirovskii, 1994] Nesterov, Y. and Nemirovskii, A. (1994). *Interior-point polynomial algorithms in convex programming*, volume 13. SIAM.

[Nesterov and Nemirovsky, 1992] Nesterov, Y. and Nemirovsky, A. (1992). Conic formulation of a convex programming problem and duality. *Optimization Methods and Software*, 1(2):95–115.

[Nesterov and Tunçel, 2016] Nesterov, Y. and Tunçel, L. (2016). Local superlinear convergence of polynomial-time interior-point methods for hyperbolicity cone optimization problems. *SIAM Journal on Optimization*, 26(1):139–170.

[Nesterov and Todd, 1997] Nesterov, Y. E. and Todd, M. J. (1997). Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations research*, 22(1):1–42.

[Nesterov and Todd, 1998] Nesterov, Y. E. and Todd, M. J. (1998). Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on optimization*, 8(2):324–364.

[Nie, 2012] Nie, J. (2012). Sum of squares methods for minimizing polynomial forms over spheres and hypersurfaces. *Front. Math. China*, 7:321–346.

[Ralph, 1994] Ralph, D. (1994). Global convergence of damped Newton's method for nonsmooth equations via the path search. *Mathematics of Operations Research*, 19(2):352–389.

[Renegar, 1988] Renegar, J. (1988). A polynomial-time algorithm, based on newton's method, for linear programming. *Mathematical programming*, 40(1):59–93.

[Renegar, 2001] Renegar, J. (2001). *A Mathematical View of Interior-point Methods in Convex Optimization*. MPS-SIAM Series on Optimization. SIAM, MPS.

[Reznick, 1978] Reznick, B. (1978). Extremal PSD forms with few terms. *Duke Math. J.*, 45(2):363–374.

[Reznick, 2000] Reznick, B. (2000). Some concrete aspects of Hilbert's 17th problem. *Contemporary Mathematics*, 253:251–272.

[Saunderson, 2019] Saunderson, J. (2019). Certifying polynomial nonnegativity via hyperbolic optimization. *SIAM J. Appl. Algebra Geom.*, 3(4):661–690.

[So, 2011] So, A. M.-C. (2011). Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems. *Math. Program.*, 129:357–382.

[Tanabe, 1988] Tanabe, K. (1988). Centered newton method for mathematical programming. In *System Modelling and Optimization: Proceedings of the 13th IFIP Conference Tokyo, Japan, August 31–September 4, 1987*, pages 197–206. Springer.

[Taylor et al., 2018] Taylor, A., Van Scoy, B., and Lessard, L. (2018). Lyapunov functions for first-order methods: Tight automated convergence guarantees. In *International Conference on Machine Learning*, pages 4897–4906. PMLR.

[Todd et al., 1998] Todd, M. J., Toh, K.-C., and Tütüncü, R. H. (1998). On the nesterov–todd direction in semidefinite programming. *SIAM Journal on Optimization*, 8(3):769–796.

[Todd and Ye, 1990] Todd, M. J. and Ye, Y. (1990). A centered projective algorithm for linear programming. *Mathematics of Operations Research*, 15(3):508–529.

[Toh et al., 1999] Toh, K.-C., Todd, M. J., and Tütüncü, R. H. (1999). Sdpt3—a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581.

[Tunçel, 2000] Tunçel, L. (2000). Potential reduction and primal-dual methods. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, pages 235–265.

[Vanderbei et al., 1986] Vanderbei, R. J., Meketon, M. S., and Freedman, B. A. (1986). A modification of karmarkar's linear programming algorithm. *Algorithmica*, 1:395–407.

[Zhang et al., 2012] Zhang, X., Qi, L., and Ye, Y. (2012). The cubic spherical optimization problems. *Math. Comput.*, 81:1513–1525.