Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Computer Science

Time-Varying Semidefinite Programming

Dissertation thesis
for the degree of Doctor of Philosophy

PhD programme:	Electrical Engineering and Information Technology
Branch of study:	Informatics - Department of Computer Science

Supervisor :	Dr. Jakub Mareček, PhD
Co-Supervisor:	Prof. Ing. Didier Henrion, PhD
Specialist Supervisor:	Dr. Vyacheslav Kungurtsev, PhD

Antonio Bellon

Prague - October 25, 2023

Supervisor:
Dr. Jakub Mareček, PhD
Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo náměstí 13
121 35 Prague 2
Czech Republic

Co-Supervisor:
Prof. Ing. Didier Henrion, PhD
LAAS-CNRS
7 avenue du Colonel Roche
314 00 Toulouse
France

Department of Control Engineering
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo náměstí 13
121 35 Prague 2
Czech Republic

Specialist Supervisor:
Dr. Vyacheslav Kungurtsev, PhD
Department of Computer Science
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo náměstí 13
121 35 Prague 2
Czech Republic

Abstract

Semidefinite programming (SDP) is a large class of linear conic optimization problems, whose variables are matrices belonging to an affine section of the positive semidefinite cone. In many applications, the input data of these convex optimization problems change as a function of time. This thesis explores time-varying semidefinite programs (TV-SDPs), SDP problems whose data and solutions depend on a time parameter.

We first study the geometry of the trajectory of solutions, defined as the set-valued map that associates to any value of the time parameter the set of optimal solutions. We propose an exhaustive description of the geometric behavior of this trajectory. As our main result, we show that along the solution trajectory only six distinct types of behaviors can be observed, and we illustrate each type with an example.

Next, we present a path-following algorithm for TV-SDP, based on tracking the solutions trajectory of a matrix factorization, known as the Burer–Monteiro factorization. The method is built on solving a sequence of linearized optimality conditions systems, which requires the introduction of a horizontal space constraint to ensure the local injectivity of the factorization. The algorithm produces a sequence of approximate solutions for the original TV-SDP problem, for which we show that they stay close to the optimal solution path if properly initialized. Numerical experiments for a time-varying Max-Cut SDP relaxation demonstrate the computational advantages of the method for tracking TV-SDPs in terms of runtime and accuracy, compared to off-the-shelf interior point algorithms.

Keywords: semidefinite programming; parametric optimization; convex optimization; linear conic optimization; continuous optimization; matrix optimization

Abstrakt

Semidefinitní programování (SDP) je rozsáhlá třída konvexních optimalizačních úloh, jejichž proměnnými jsou matice patřící do afinního podprostoru positivně-semidefinitního kužele. V mnoha aplikacích se vstupní data těchto konvexních optimalizačních problémů mění jako funkce času. Tato práce zkoumá v čase proměnné semidefinitní programy (TV-SDP), jejichž data a řešení závisí na parametru času.

Nejprve studujeme geometrii trajektorie řešení, definovanou jako funkcí s oborem hodnot na množinách, která k libovolné hodnotě časového parametru přiřazuje množinu optimálních řešení. Představujeme dobrou charakterizaci geometrického chování této trajektorie. Jako náš hlavní výsledek ukazujeme, že podél trajektorie řešení lze pozorovat pouze šest různých typů chování, a každý typ ilustrujeme na příkladu.

Dále představujeme algoritmus pro sledování trajektorie řešení TV-SDP, který je založen na sledování trajektorie řešení maticové faktorizace, známé jako Burerova–Monteirova faktorizace. Algoritmus je postaven na řešení posloupnosti linearizovaných systémů podmínek optimality, což vyžaduje zavedení omezení na tzv. horizontální podprostor, které zajistí lokální injektivitu faktorizace. Algoritmus vytváří posloupnost přibližných řešení původního problému TV-SDP, u nichž ukazujeme, že při správné inicializaci zůstávají blízko optimální trajektorie řešení. Numerické experimenty pro v čase proměnnou relaxaci problému bisekce ukazují, že algoritmus je z hlediska doby běhu a přesnosti vylepšením oproti standardním algoritmům založeným na tzv. metodě vnitřního bodu.

Klíčová slova: semidefinitní programování; parametrická optimalizace; konvexní optimalizace; lineární konická optimalizace; spojitá optimalizace; maticová optimalizace

Alla mia famiglia

Acknowledgements

This research has been supported by the OP RDE funded project *Research Center for Informatics* (CZ.02.1.01/0.0/0.0/16_019/0000765) and by the Czech Science Foundation (contract 22-15524S).

Program

List of Publications

- Bellon, A., Henrion, D., Kungurtsev, V., and Mareček, J., “Parametric semidefinite programming: geometry of the trajectory of solutions", to appear in *Mathematics of Operations Research*, preprint in arXiv:2104.05445, 2021.
- Bellon, A., Dressler, M., Kungurtsev, V., Mareček, J., and Uschmajew, A., “Time-varying semidefinite programming: path following a Burer–Monteiro factorization", to appear in *SIAM Journal on Optimization*, preprint in arXiv:2210.08387, 2022.
- Bellon, A., Liu, J., Mareček, J., Simonetto, A., and Takáč, M. “Optimal Power Flow Pursuit in the Alternating Current Model", preprint in arXiv:2211.02939, 2022.

List of Tables

List of Figures

List of Acronyms

KKT Karush–Kuhn–Tucker

LICQ Linear Independence Constraint Qualification

LMI Linear Matrix Inequality

LP Linear Programming

OPF Optimal Power Flow

POP Polynomial Optimization

SDP Semidefinite Programming

SOCP Second Order Cone Programming

TV-LP Time-Varying Linear Programming

TV-MCR Time-Varying Max Cut Relaxation

TV-POP Time-Varying Polynomial Optimization

TV-SDP Time-Varying Semidefinite Programming

TV-SOCP Time-Varying Second Order Cone Programming

List of Symbols

Scalars

n, m, r, s, κ	natural numbers
i, j, k	natural indexes
$\alpha, \beta, \gamma, \delta, \varepsilon, \eta, \tau$	real numbers
t	time parameter
$p^\star/d^\star$	primal/dual optimal values

Vectors

a, b, c, x, y, z	real finite dimensional vectors
y, μ, ν	vectors of Lagrangian multipliers

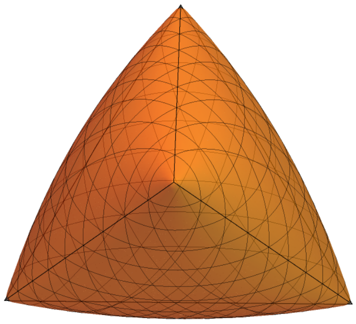
Sets, spaces, and manifolds

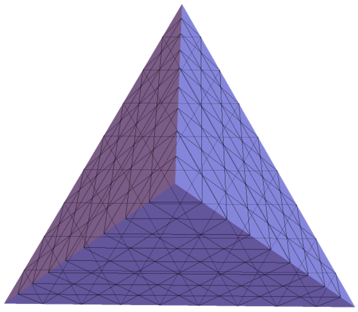
$[n]$	set of first n natural numbers $\{1, \ldots, n\}$
$\mathbb{N}$	natural numbers
$(0, \tau)/[0, \tau]$	real open/closed intervals
$\mathbb{R}$	real numbers
$\overline{\mathbb{R}}$	extended real numbers $\mathbb{R} \cup \{-\infty, \infty\}$
$\mathbb{R}^n$	space of real n-dimensional vectors
$\mathbb{R}^n_+$	n-dimensional non-negative orthant
$\mathbb{L}^n$	Lorentz cone in $\mathbb{R}^n$
$\mathbb{R}^{n\times m}$	space of real $n \times m$ matrices
$\mathbb{R}^{n\times r}_*$	set of $n \times r$ matrices with full column rank
$\mathbb{C}$	complex numbers
$\mathbb{C}^n$	space of complex n-dimensional vectors
$\mathbb{C}^{n\times m}$	space of complex $n \times m$ matrices

$\mathbb{S}^n$	space of real symmetric $n \times n$ matrices
$\mathbb{S}^n_{skew}$	space of real skew-symmetric $n \times n$ matrices
$\mathbb{S}^n_+$	cone of positive semidefinite $n \times n$ matrices
$\mathbb{E}, \mathbb{F}$	Euclidean spaces
$\mathcal{B}_r(x)$	ball of radius r centered on x
$\mathcal{O}_r$	group of $r \times r$ orthogonal matrices
$\mathcal{M}_r$	manifold of fixed rank-r symmetric matrices
$\mathcal{M}_r^+$	manifold of fixed rank-r positive semidefinite matrices.
$\mathcal{T}_X$	tangent space at X
$\mathcal{H}_Y$	horizontal space at Y
$\mathcal{P}/\mathcal{D}/\mathcal{P}^\star/\mathcal{D}^\star$	primal/dual feasible/optimal sets
$\dim \mathcal{S}$	dimension of a linear/affine space or manifold $\mathcal{S}$

Matrices and operators

A, B, C, X, Y, Z	matrices
A^T, A^{-1}	transpose and inverse of a matrix A
I_n	$n \times n$ identity matrix
$\mathrm{diag}(v)$	diagonal matrix with a given vector v on the diagonal
$\succeq$	Lowner partial order on square matrices
$\mathcal{A}, \mathcal{A}^*$	linear operator and its adjoint
$\mathcal{A}^{-1}, \mathcal{A}^\dagger$	inverse and pseudo-inverse of a linear operator $\mathcal{A}$
$\ker A/\mathcal{A}$	null space of a matrix A or a linear operator $\mathcal{A}$
$\mathrm{im} A/\mathcal{A}$	image of a matrix A or a linear operator $\mathcal{A}$
$\mathrm{span} A$	linear space spanned by the columns of a matrix A
$\mathrm{rank} A$	dimension of $\mathrm{span} A$
$\det A$	determinant of a matrix A
$\mathrm{trace} A$	trace of a matrix A
$\sigma_i(A)$	i-th singular value of a matrix A in decreasing order
$\lambda_i(A)$	i-th eigenvalue of a square matrix A in decreasing order
$\langle \cdot, \cdot \rangle$	Euclidean or Frobenius inner product
$\|\cdot\| / \|\cdot\|_F$	Euclidean/Frobenius norm
$\otimes_s$	symmetric Kronecker product





Chapter 1

Ouverture

Who doesn't love opera? And who doesn't love mathematics? After all, they both offer great and immortal stories. Regrettably, far too many do not appreciate these two great products of the human intellect. This work is not for them. To be fair, if the reader enjoys opera but not mathematics, this work is not for them either. On the other hand, if the reader loves mathematics but not opera, they can surely keep reading. We do hope that the reader is fond of both and that will enjoy this operatic thesis. Please sit comfortably, turn off the phone, and let the maths begin...

"Oh che bella introduzione vi sarebbe da cavar!"

Prosdocimo in Il Turco in Italia (Act I, Scene 1)

Semidefinite programming (SDP)[1] is a class of convex constrained optimization problems. More precisely, as we will see in Chapter 2, it is the set of linear conic optimization problems over the cone of positive semidefinite matrices. To put it into a perhaps more familiar perspective for the average reader, SDP can be thought of as a generalization of linear programming (LP). This latter is probably the most well-known and studied class of problems in the field of mathematical optimization, and it can now be regarded as a reliable technological tool. Its story is indeed a successful one: its theory is today well understood, more and more efficient algorithms have been developed, and its applications are spread around many fields of mathematics, engineering, and computer science.

While the origins of LP date back to the 1940s, SDP is a much younger field, as it started to gain the attention of the optimization community only in the late 1980s. Since then, many researchers from very different fields have been attracted to this class of problems, including experts in convex programming, linear algebra, numerical

[1] The acronym SDP can also denote a "semidefinite program". Its use will be clear from the context.

optimization, combinatorial optimization, polynomial optimization, control theory, and statistics, who were interested in the theoretical, numerical, and application aspects of these problems. This research activity has greatly increased since the discovery of important applications in combinatorial optimization [1, 2] and control theory [3, 4], as well as the development of efficient interior point methods.

In particular, the success of SDP was boosted by the discovery that semidefinite relaxations are available for fundamental problems such as the Max-Cut problem, for which the successful Goemans–Williamson approximation algorithm [5] is available, or coloring problems [6].

Meanwhile, the development of efficient polynomial-time interior point algorithms [7, 8, 9, 10], [11, Chapter 9] for solving SDPs drastically contributed to making SDP an interesting and powerful tool in conic optimization. These methods solve a convex optimization problem by generating a sequence of points lying in the interior of the feasible set and converging to a boundary point that corresponds to an approximate optimal solution of any desired accuracy. The generation of this sequence is based on the application of Newton's method to find a solution to systems of perturbed optimality conditions. As a perturbation parameter is tuned, the solution of the systems converges to an optimal solution, following a smooth curve, the so-called *central-path*. These algorithms are second-order methods, so that they require to iteratively compute the Hessian of a function. First-order method, only requiring gradient computations, are also available, but they are typically slower.

More recently, the importance of SDP was established by the development of Lasserre's Moment-Sum-of-Squares hierarchy [12, 13] for solving a polynomial optimization problem (POP). An approximate solution for the latter is found by solving a sequence of SDP problems of increasing size, which are convex relaxations of an equivalent formulation of the original POP as an infinite dimensional linear problem in the space of finite measures. Many problems in control theory, such as checking Lyapunov stability of dynamical systems, can be reduced to solving polynomial equations, polynomial inequalities, or polynomial differential equations, and they can hence be approximately solved by the Moment-Sum-of-Squares hierarchy [14]. A survey of the state of the art in the areas of SDP, conic optimization, and polynomial optimization is given in [15].

Other notable and interesting applications of SDP appear in machine learning [16, 3], quantum information [17], conformal fields theory [18], finance [19], as well as in estimation [20], graph realization [21], and matrix completion [22] problems.

Time-varying semidefinite programming (*Allegro*)

A semidefinite program is an optimization problem that can be expressed in the form

$$\begin{aligned} p^\star = \ & \inf_{X \in \mathbb{S}^n} \ \langle C, X \rangle \\ & \text{s.t.} \ \ \mathcal{A}(X) = b, \\ & \qquad X \succeq 0, \end{aligned} \tag{P}$$

which we will refer to as the *primal* problem. The goal of (P) is to find a symmetric matrix X minimizing the objective function over a feasible region defined by a set of linear equations and a positive semidefiniteness constraint, finding the optimal value $p^\star \in \overline{\mathbb{R}}$. The objective function is linear and usually defined as the Frobenius inner product

$$\langle C, X \rangle := \operatorname{trace}(C^T X) = \sum_{i,j=1}^{n} C_{i,j} X_{i,j} \tag{1.1}$$

between a matrix C and the matrix variable X, both belonging to $\mathbb{S}^n$, the space of symmetric $n \times n$ matrices of with real entries. Indeed, $\mathbb{S}^n$ equipped with the inner product (1.1) forms a Euclidean space. The notation $\mathcal{A}(X) = b$ models a number m of linear equations that X must satisfy, $\langle A_i, X \rangle = b_i$ for $i \in [m]$, where $A_i \in \mathbb{S}^n$ are given matrices and b_i are given scalars, thus defining a linear operator $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ between Euclidean spaces. Furthermore, the matrix variable X must belong to the set of positive semidefinte matrices

$$\mathbb{S}^n_+ := \{X \in \mathbb{S}^n \mid v^T X v \geq 0, \ \forall v \in \mathbb{R}^n\},$$

which, as discussed in the first section of the next chapter, forms a proper cone. The notation $X \succeq 0$ is a shortcut for $X \in \mathbb{S}^n_+$, and the symbol $\succeq$ denotes the Loewner partial order on square matrices:

$$A \succeq B \iff A - B \in \mathbb{S}^n_+.$$

The feasible region is hence the intersection of the semidefinite cone with an affine subspace described by linear equations:

$$\mathcal{P} := \{X \in \mathbb{S}^n \mid \mathcal{A}(X) = b, X \succeq 0\}.$$

Such a set, often referred to as a *spectrahedron*, is convex and basic closed semialgebraic, as it can be expressed as the intersection of n polynomial inequalities. These can be derived by imposing that the n coefficients of the characteristic polynomial $\pi_X(\lambda) = \det(\lambda I_n - X)$ of a matrix X are such that the solutions of $\pi_X(\lambda) = 0$, that is, the eigenvalues of X, are all non-negative (see, e.g., Corollary 7.2.4 in [23]). Thus, problem (P) is a convex optimization problem whose feasible region is a spectrahedron. If $\mathcal{P}$ is empty, we say that the problem (P) is *infeasible*, and we write $p^\star = \infty$. Otherwise, we say that the problem is *feasible*. Finally, if $p^\star = -\infty$ we say that the problem is *unbounded*. We refer to any $X \in \mathcal{P}$ as a *feasible solution* or a *feasible point*. If, furthermore, $\langle C, X \rangle = p^\star$ we say that X is an *optimal solution* or an *optimal point*. Note that, in principle, the infimum in (P) might not be attained. However, under standard constraints qualification assumptions that we review Chapter 2, the infimum is indeed a minimum.

The subject of this thesis is time-varying semidefinite programming (TV-SDP). A TV-SDP is a parametric SDP problem that can be expressed in the form

$$\begin{aligned} p_t^\star = \inf_{X \in \mathbb{S}^n} \ & \langle C_t, X \rangle \\ \text{s.t.} \ & \mathcal{A}_t(X) = b_t, \\ & X \succeq 0. \end{aligned} \tag{P_t}$$

The objects defining (P_t) are the same as those defining (P), with the only difference being the dependence of the problem data $\mathcal{A}, b, C$ on a parameter t, which we think as a time parameter, varying on a given real interval $[0, \tau]$. In this time-varying setting, the goal is to find a curve $t \mapsto X_t$ in $\mathbb{S}^n$ such that $X = X_t$ is an optimal solution for (P_t) at each time point $t \in [0, \tau]$. Numerically, this translates in finding a sequence of matrices $\{X_k\}_{k \in [\kappa]}$ such that each X_k is an (approximate) optimal solution to (P_{t_k}), where $\{t_k\}_{k \in [\kappa]}$ is a partition of size $\kappa \in \mathbb{N}$ of the parametrization interval $[0, \tau]$. We adopt the former point of view in Chapter 3, where we study the geometric properties of the trajectory of solutions, while in Chapter 4 we address the problem from the latter perspective, providing an algorithm that produces a sequence of solutions $\{X_k\}_{k \in [\kappa]}$.

Time-dependent problems leading to TV-SDPs occur in various applications, such as optimal power flow problems in power systems [24], state estimation problems in quantum systems [17], modeling of energy economic problems [25], job-shop scheduling problems [26], as well as problems arising in signal processing, queueing theory [27] or aircraft engineering [28].

In order to illustrate the modeling power of TV-SDP, we briefly describe three notable examples. The optimal power flow (OPF) is a non-convex and generally hard to solve problem in power systems, where it is critical for maintaining secure and economic operations. First proposed by Carpentier [29], it has become increasingly important as electric power system operations grew in complexity. Its objective is usually minimizing the sum of the costs of power generators under the conditions of operating constraints. The OPF problem has been extensively studied in the literature and numerous algorithms have been proposed for solving this non-convex problem. Among these, semidefinite programming relaxations are very successful. For a highly cited example, see [24], where a semidefinite program, which is the dual of an equivalent form of the OPF problem, was proposed. There, a globally optimal solution to the OPF problem can be retrieved from a solution of this convex dual problem whenever the relaxation gap is zero. In real world instances of OPF problem, data are clearly time-dependent, as physical and technical conditions of the network change during the day. Hence TV-SDP appears as a natural way to address the OPF problem in its time-dependent version (see section II.A in [30]).

In quantum systems, one would like to update the estimate of the state of a quantum system as it changes under the influence of a noise process. Consider, for example, the problem where many copies of an unknown n-qubit state ρ are given. A measuring process is then defined, where at each stage only some copies of ρ are measured using a known measurement E_1. At the next stage, other copies of ρ are measured using another measurement E_2, and so on. At each stage, a current hypothesis ω_t about the state is generated using the outcomes of the previous measurements. Using techniques from online convex optimization, Aaronson et al. [17] showed that it is possible to follow this process in a way that guarantees that $|\operatorname{trace}(E_i\omega_t)-\operatorname{trace}(E_i\rho)|$, the prediction error for the next measurement, can be bounded both from below and above. Here, the key observation is that an n-qubit quantum state ρ is an element of the set of positive semidefinite complex matrices of dimension $2^n \times 2^n$ and unitary trace. Following this observation, the problem can be formulated as a TV-SDP.

In computer vision, there are also interesting applications of TV-SDPs. One is in the background subtraction problem, where the goal is to differentiate between a slow-moving background of a video sequence and the foreground, which is composed of objects moving faster than the background. This problem can be expressed as the problem of minimizing the rank of a matrix whose rows represent a subset of subsequent frames of a given video sequence [31]. Since the nuclear norm is the convex envelope of the rank function on the unit matrix ball [32], one can use it as an approximation of the rank. Furthermore, the nuclear norm is semidefinite-representable [33], so that one can cast the background subtraction problem into a TV-SDP.

Background and previous work (*Andante*)

Semidefinite programming is today a well established field of mathematical optimization, so that various textbooks and extensive surveys are available on the topic [11, 34, 35, 36, 37]. The fundamental properties of SDP (such as duality, strict feasibility, uniqueness of the solution, strict complementarity, non-degeneracy) and its prerequisites are widely understood; see for example [38], where Alizadeh, Haeberly and Overton clarified the relation between uniqueness of the solution, non-degeneracy, and strict complementarity. Likewise, the geometry of SDP, that is, spectrahedral geometry, has been studied in depth. Pataki offers in [11, Chapter 3] an excellent overview, as well as Ramana and Goldman in [39]. Non-linear SDPs were also considered in the literature, see for example [40].

On the other hand, time-varying optimization has proven to be relevant in many application (see, e.g., the survey paper [41]). TV-SDPs can be seen as a generalization of time-varying linear programming (TV-LP) problems. In the context of parametric optimization, these latter have been extensively studied [42, 43, 44, 45, 46], while Bellman [47] was the first to study them in relation to so-called bottleneck problems in multistage linear production economic processes in the context of dynamic programming [48]. Since then, a large body of literature has been devoted to studying TV-LP with and without additional assumptions. However, the generalization of this idea to other classes of optimization problems has only recently been considered, often in the context of sensitivity analysis [49, 50].

Goldfarb and Scheinberg were the first to consider in [51] parametric SDPs. There, the objective function depends linearly on a scalar parameter. Studying the properties of the optimal value as a function of the parameter, they extended the concept of optimal partition from LP to SDP.

More recently, Ahmadi and El Khadir [52] considered time-varying SDPs. In contrast to our setting, in their work they require the data to vary polynomially with time, proving that it is possible to restrict to consider solutions whose entries also vary polynomially with time without changing the optimal value.

Following the pioneering contribution of Goldfarb and Scheinberg [51], a number of important papers appeared recently:

- Al-Salih and Bohner [53] studied LP on time scales, which allows for the mixing of difference and differential operators in a broad class of extensions of LP models using the notion of time scales.

- Wang, Zhang, and Yao [54] studied a wide family of parametric optimization problems, which are known as separated continuous conic programming, and developed a strong duality theory for these. They proposed a polynomial-time approximation algorithm that solves such problems to any required accuracy. Their framework is generalized to SDP in [52].

- Nayakkankuppam and Overton [55] studied the sensitivity of SDPs, analyzing the effect on the solution of small perturbations on the problem data. They derived an explicit bound on the change in the solution in a primal-dual framework, quantifying the size of permissible perturbations.

- Sekiguchi and Waki [4] studied sensitivity of an SDP under perturbations in a more general framework than in [51], showing that when the coefficient matrices are perturbed, the optimal values can change discontinuously and illustrate this in concrete examples. To some extent, we move in parallel to their results, studying the behavior of optimal solutions of a TV-SDP as function of the time parameter, while they focus on the optimal value instead.

- Mohammad-Nezhad [56], together with Terlaky [57], Haunstein, and Tang [58], are the closest to our work in spirit. Their dependence of the problem on the data is assumed to be linear, which is a more restrictive assumption than the one we use. In part, we build upon their theoretical results, but instead of employing the concepts of non-linearity intervals, invariancy intervals, and transition points, we use a purely set-valued analysis approach.

- Ahmadi and El Khadir [59] and [52] are the closest to our work in name. They studied the setting where the data vary with known polynomials of the parameter and showed that under a strict feasibility assumption, restricting the solutions to be polynomial functions of the parameter does not change the optimal value of the TV-SDP. They also provided a sequence of SDP problems that give upper bounds on the optimal value of a TV-SDP converging to the optimal value. In contrast, as discussed in the next paragraph, we use a different setting, where for the first of our main results, Theorem 3.3 we only assume continuity of the map from the parameter to the problem data, while we make a stronger polynomial dependence assumption for Theorem 3.4. Moreover, we provide a complete geometric characterization of the solutions trajectory.

It is worth at this point to remark that there are mainly two different types of TV-SDPs considered in the literature. The first type is the one considered by the aforementioned [59] and [52]. There, the constraints at a given time point are linked to to the solutions at the previous times via kernel terms. In this case, the solutions are thought as measurable functions, which are required to satisfy the constraints only on a set of times that is the complement of a measure-zero set, i.e. *almost everywhere*. Instead, we consider the easier case where constraints are independent through time. This approach, considered by [51, 60] and [58], assumes that the coefficients of the SDP simply depend on a parameter, and looks for solutions of the problem at each value of the parameter.

In our framework, the adjective time-varying hence simply refer to the presence of a univariate parameter, which in applications often coincides with time.

Contributions (*Rondò*)

Our primary goal is to investigate the properties of the trajectories of solutions to TV-SDP, on the one hand to increase the knowledge in an interesting and relatively novel area of research; on the other hand, to set the theoretical bases to design efficient algorithms for TV-SDP with guarantees on their performance. The main contents of the thesis are exposed in Chapters 3 and 4, and follow closely two papers: *Parametric SDP: geometry of the trajectory of solutions* [61] and *TV-SDP: path following a Burer–Monteiro factorization* [62].

In particular, we want to characterize points of the trajectory of solutions to TV-SDPs according to the local behavior of the trajectory at the point considered. This research objective was inspired by the textbook *Parametric Optimization: Singularities, Pathfollowing and Jumps* [63], where a classification of solutions to univariate parametric nonlinear constrained optimization problems is proposed. As a result, we define six different types of points, according to the local behavior of the trajectory of the solutions at that point. We then prove a classification theorem, which states, under general and standard assumptions, that only the types of points that we defined can actually appear. Furthermore, under the technical assumption of the existence of a generic non-singular time (see Definition 3.2), we show that only three of the six types of points that we defined can actually appear. This is the subject of Chapter 3.

The classification of points in different types is based on the geometry of the trajectory of solutions over a given time parametrization interval. Before a given time, we assume that the trajectory is regular and follows a continuous curve. Then, at the time of interest, we distinguish between the following situations:

- **Regular point:** the trajectory is single-valued and differentiable;
- **Non-differentiable point:** the trajectory is single-valued but not differentiable;
- **Discontinuous isolated multiple point:** a loss of continuity causes a loss of uniqueness of the solution, implying a multiple-valued solution. After the point, uniqueness is restored, and hence the loss of uniqueness is isolated;
- **Discontinuous non-isolated multiple point:** a loss of continuity causes a loss of uniqueness of the solution, implying a multiple-valued solution. After the point, uniqueness is not restored hence the loss of uniqueness is not isolated;
- **Continuous bifurcation point:** the trajectory splits into several distinct branches. This results in a loss of uniqueness which still preserves continuity;
- **Irregular accumulation point:** accumulation point of a set made of either bifurcation points or discontinuous isolated multiple points.

We believe that a first contribution of our research is precisely the definition of these types of points. The main results presented in Chapter 3 are Theorem 3.3 and 3.4, which we informally state here.

Informal statement of Theorem 3.3 Under assumptions of linear independence constraint qualification (LICQ), existence of strictly feasibile point and continuity of the data with respect to time, the trajectory can only be comprised of points of the six types described above.

Informal statement of Theorem 3.4 Under the same assumptions of Theorem 3.3, suppose that the problem data are polynomial functions of time and that there exists a generic non-singular time (see Def. 3.2). Then the trajectory is comprised of only regular points, non-differentiable points, or isolated multiple points. In other words, non-isolated discontinuous multiple points, bifurcation points, and irregular accumulation points cannot appear. Furthermore, the number of non-differentiable points, or isolated multiple points is *finite*.

Notice that while we only assume continuity of the data for Theorem 3.3, we need stronger regularity assumptions to guarantee the validity of Theorem 3.4. The two results are visually summarized in the next page by Table 1.1.

As mentioned before, other than interesting for a purely theoretical study, these results were useful for the subsequent development of a path-following algorithm to track the

Problem assumptions	**Type of points**
TV-SDP with LICQ, polynomial data, strict feasibility, and a generic *non-singular* time	Regular points Non-differentiable points Discontinuous isolated multiple points
TV-SDP with LICQ, continuous data, strict feasibility, without a generic *non-singular* time	Regular points Non-differentiable points Discontinuous isolated multiple points Discontinuous non-isolated multiple points Continuous bifurcation points Irregular accumulation points

Table 1.1: Assumptions on TV-SDP and associated possible type of points

trajectory of solutions to TV-SDP, similarly to what has been developed in [63, 64]. In classical predictor-corrector methods, a predictor step for approximating the directional derivative of the solution with respect to a small change in the time parameter is applied, together with a correction step that moves from the current approximate solution closer to the next solution at the new time point. The method that we propose incorporates these two steps in a single Newton step applied to the first-order optimality conditions.

Indeed, if one wants to apply a path-following strategy, the trajectory of solutions needs to be defined by a regular (smooth) curve. Thanks to our previous analysis, in particular to Theorem 3.4, we know that, under the condition of the existence of a generic non-singular time, the trajectory of solutions is made of regular points, except for a finite number of loss of differentiability and discontinuous isolated multiple points. This justifies and motivates the development of a procedure that can track a trajectory when this is only made of regular points, hence describing a smooth curve. The analysis of the possible irregular behaviors that the trajectory of solutions can be affected from has the further advantage of providing algebraic conditions that in principle can be numerically monitored to check whether the trajectory is approaching a point where regularity is loss.

A limiting factor in solving both stationary and time-dependent SDPs is computational complexity when n is large. A common solution to this obstacle is the Burer–Monteiro approach, as presented in the seminal work [65, 66]. In this approach, a low-rank Cholesky factorization $X = YY^T$ of the solution is assumed with $Y \in \mathbb{R}^{n\times r}$ and r potentially much smaller than n. This leads to the following factorized version of (P)

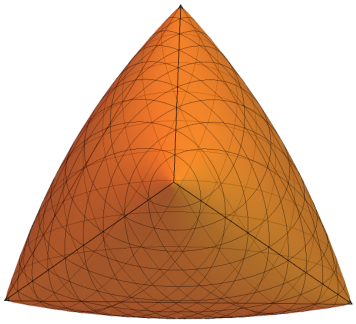
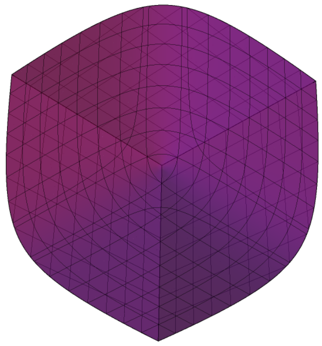
$$\begin{aligned} \min_{Y\in\mathbb{R}^{n\times r}} \quad & \langle C_t, YY^T\rangle \\ \text{s.t.} \quad & \mathcal{A}_t(YY^T) = b_t, \end{aligned} \tag{Q_t}$$

which is a quadratically constrained quadratic problem, that is in general non-convex.

Moreover, theoretically it may happen that local optimization methods converge to a critical point that is not globally optimal [67], although in practice the method usually shows very good performance [65, 68, 69]. This approach has been carefully studied in the optimization literature, e.g. in terms of algorithms [68], quality of the optimal value [70, 71], and (global) recovery guarantees [72, 73, 74].

In Chapter 4, we show how to combine the Burer–Monteiro factorization with path-following methods, developing a practical algorithm for approximating the solution of (Q_t), and consequently of (P_t), over time. As a main obstacle to this strategy, to apply such methods, one needs to address the issue that the solutions of (Q_t) are never isolated. This is due to the non-uniqueness of the Burer–Monteiro factorization, which is caused by the orthogonal invariance of the solutions to (Q_t): if Y_t is a solution for a given t, then so is $Y_t Q$ for any orthogonal matrix Q, since $(Y_t Q)(Y_t Q)^T = Y_t Y_t^T$. To handle this problem, we apply a well-known technique by restricting the solutions to a so-called *horizontal space* at every time step, enforcing the isolation of solutions to (Q_t).

Exploiting these ideas, which are described in detail in the first section of Chapter 4, in the subsequent section we fully develop this path-following method in Algorithm 1, contextually showing that the restricted problem satisfies second-order sufficiency conditions for optimality. This is able to generate a sequence $\{(X_k, Z_k)\}_{k\in[\kappa]}$ of primal-dual optimal solutions to ($\mathrm{P}_t, \mathrm{D}_t$) as accurate as required. We also provide a result which proves that the tracking is performed with bounded accuracy. In the last section of the chapter, we test the method against a time-varying version of the SPD relaxation for the Max-Cut problem. We illustrate numerical experiments conducted using a Python implementation, of the algorithm, showing that the path-following procedure that we propose exhibit promising computational benefits, in terms of both accuracy and runtime.

Chapter 2

Act I Preliminaries

"Was anders ist, das lerne nun auch!"

The Wonderer/Wotan in Siegfried (Act II, Scene 1)

In this chapter, we expose the theoretical basics and the technical tools necessary for the discussion in the next chapters of our main results: a classification theorem and a path-following algorithm. After showing how SDP arises in the context of linear conic optimization, we review the standard properties of SDP. We then survey continuity properties of the optimal and feasible sets of TV-SDP, considered as set-valued maps, in terms of inner and outer semi-continuity and Painlevé–Kuratowski continuity.

Scene 1 Conic optimization and duality

Where we get to know our heroes, the SDP prince and its twin, the dual prince, and we discover that they are the rightful heirs of the noble family of Linear Conic Optimization problems led by none other than Her Majesty the Queen of Cones, mother of the SDP princes. We are then given a family portrait in which our two protagonists proudly stand out among their illustrious cousins, the powerful dukes of LP and the wealthy earls of SOCP...

In this section, we define linear conic optimization and review the rather elegant theory of conic duality. Let us recall that a *Hilbert space* $\mathbb{H}$ is a complete vector space equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}} : \mathbb{H} \times \mathbb{H} \to \mathbb{R}$. This induces a norm on $\mathbb{H}$, defined by the relation $\|x\|_{\mathbb{H}}^2 = \langle x, x \rangle_{\mathbb{H}}$. Finite dimensional Hilbert spaces are called *Euclidean spaces*, and we denote them by $\mathbb{E}$. The space of real n-dimensional vectors $\mathbb{R}^n$, endowed with the usual scalar product is probably the most familiar example of a Euclidean space, while in this dissertation we are mainly interested in the Euclidean space of real $n \times n$ matrices $\mathbb{S}^n$ equipped with the Frobenius inner product.

Example 2.1 (Relevant Euclidean spaces with the relative inner products)**.**

1. $\mathbb{R}^n$ with $\langle x, y\rangle_{\mathbb{R}^n} = x^T y = \sum_{i=1}^{n} x_i y_i,$
2. $\mathbb{S}^n$ with $\langle X, Y\rangle_{\mathbb{S}^n} = \operatorname{trace}(X^T Y) = \sum_{i,j=1}^{n} X_{i,j} Y_{i,j}.$

In the following chapters, we will omit the space in which the scalar product is defined, as this will be clear from the context. Among the interesting subsets of Euclidean spaces, we focus on proper cones.

Definition 2.1 (Cones)**.** Given a Euclidean space $\mathbb{E}$, a *cone* is a subset $\mathcal{K} \subseteq \mathbb{E}$ such that for every $x \in \mathcal{K}$ and $\alpha \geq 0$ we have $\alpha x \in \mathcal{K}$. A *convex cone* is a set $\mathcal{K}$ such that for any $x, y \in \mathcal{K}$ and $\alpha, \beta \geq 0$, we have $\alpha x + \beta y \in \mathcal{K}$. A set is *pointed* if it contains no lines: $x, -x \in \mathcal{K} \implies x = 0$. A *proper cone* is a convex cone which is closed, with non-empty interior and pointed.

Example 2.2 (Relevant cones)**.**

1. $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\} = \{x \in \mathbb{R}^n \mid x_i \geq 0,\ \forall i \in [n]\},$
2. $\mathbb{L}^n = \{(x_0, x) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x\| \leq x_0\}$
3. $\mathbb{S}^n_+ = \{X \in \mathbb{S}^n \mid X \succeq 0\} = \{X \in \mathbb{S}^n \mid v^T X v \geq 0,\ \forall v \in \mathbb{R}^n\},$

For example, 1. the non-negative orthant of $\mathbb{R}^n$, 2. the second-order (or Lorentz) cone in $\mathbb{R}^n$, and 3. the set of positive semidefinite matrices of $\mathbb{S}^n$, are all proper cones.

For any proper cone, we define its dual cone.

Definition 2.2 (Dual cones)**.** Given a convex cone $\mathcal{K}$ its *dual cone* is

$$\mathcal{K}^* = \{y \in \mathbb{E} \mid \langle x, y\rangle_{\mathbb{E}} \geq 0 \text{ for all } x \in \mathcal{K}\}.$$

It is easy to see that the cones defined in Example 2.2 are all self-dual.

Example 2.3 (Dual cones of relevant cones)**.**

$$1. \quad (\mathbb{R}^n_+)^* = \mathbb{R}^n_+, \qquad 2. \quad (\mathbb{L}^n)^* = \mathbb{L}^n, \qquad 3. \quad (\mathbb{S}^n_+)^* = \mathbb{S}^n_+.$$

A linear conic programming is the class of convex optimization problems where one wants to optimize a linear objective function over the intersection of a convex cone $\mathcal{K}$ with an affine subspace of an Euclidean space $\mathbb{E}$:

$$\begin{aligned} p^\star = \quad & \inf_{x\in\mathbb{E}} \ \langle c, x\rangle_{\mathbb{E}} \\ & \text{s.t.} \ \ \mathcal{A}(x) = b, \\ & \qquad x \in \mathcal{K}. \end{aligned} \tag{C_p}$$

The goal of (C_p) is to find the infimum of the objective function among feasible vectors $x \in \mathbb{E}$, given the data of a linear operator $\mathcal{A} : \mathbb{E} \to \mathbb{F}$, a vector $b \in \mathbb{F}$, and a vector $c \in \mathbb{E}$.

In standard optimization terminology, the above problem is called the *primal problem*, and $p^\star$ denotes its infimum. For a primal conic problem its *dual problem* is defined as

$$\begin{aligned} d^\star = \quad & \sup_{y\in\mathbb{F}, z\in\mathbb{E}} \ \langle b, y\rangle_{\mathbb{F}} \\ & \text{s.t.} \ \ \mathcal{A}^*(y) + z = c, \\ & \qquad z \in \mathcal{K}^*, \end{aligned} \tag{C_d}$$

where $\mathcal{A}^* : \mathbb{F} \to \mathbb{E}$ is the adjoint operator of $\mathcal{A}$, defined so that $\langle \mathcal{A}(x), y\rangle_{\mathbb{F}} = \langle x, \mathcal{A}^*(y)\rangle_{\mathbb{E}}$ holds for all $x \in \mathbb{E}, y \in \mathbb{F}$, and $d^\star$ indicates the supremum.

Table 2.1 below shows the primal and dual form of three notable examples of linear conic programming: Linear Programming (LP) is linear optimization on the non-negative orthant, the Second Order Cone Programming (SOCP) is linear optimization on the Lorentz cone, while Semidefinite Programming (SDP) is linear optimization on the cone of positive semidefinite matrices.

Directly from the definitions of the primal problem (C_p) and its dual (C_d) it follows

Proposition 2.1 (Weak duality)**.**

$$d^\star \le p^\star \tag{2.1}$$

Proof. For optimal, hence feasible x and y we have

$$\langle c, x\rangle_{\mathbb{B}} - \langle b, y\rangle_{\mathbb{D}} = \langle c, x\rangle_{\mathbb{B}} - \langle \mathcal{A}(x), y\rangle_{\mathbb{D}} = \langle c, x\rangle_{\mathbb{B}} - \langle x, \mathcal{A}^*(y)\rangle_{\mathbb{B}} = \langle x, c - \mathcal{A}^*(y)\rangle_{\mathbb{B}} = \langle x, z\rangle_{\mathbb{B}} \ \ge \ 0.$$

$\square$

	Primal	**Dual**	**Data**
1. **LP**	$\inf_{x\in\mathbb{R}^n} c^T x$ s.t. $Ax=b$, $x\geq 0$	$\sup_{y\in\mathbb{R}^m, z\in\mathbb{R}^n} b^T y$ s.t. $A^T y + z = c$, $z \geq 0$	$A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$, $c\in\mathbb{R}^n$
2. **SOCP**	$\inf_{(x_0,x)\in\mathbb{R}\times\mathbb{R}^{n-1}} c_0x_0 + c^T x$ s.t. $a_0x_0 + Ax = b$, $\|x\| \leq x_0$	$\sup_{y\in\mathbb{R}^m, (z_0,z)\in\mathbb{R}\times\mathbb{R}^{n-1}} b^T y$ s.t. $a_0^T y + z_0 = c_0$, $A^T y + z = c$, $\|z\| \leq z_0$	$a_0\in\mathbb{R}^m$, $A\in\mathbb{R}^{(n-1)\times m}$, $b\in\mathbb{R}^m$, $c_0\in\mathbb{R}$, $c\in\mathbb{R}^{n-1}$
3. **SDP**	$\inf_{X\in\mathbb{S}^n} \operatorname{trace}(C^T X)$ s.t. $\operatorname{trace}(A_i^T X) = b_i, \quad i\in[m]$, $X\succeq 0$	$\sup_{y\in\mathbb{R}^m, Z\in\mathbb{S}^n} b^T y$ s.t. $\sum_{i=1}^m A_i y_i + Z = C$, $Z\succeq 0$	$A_i\in\mathbb{S}^n$, $b\in\mathbb{R}^m$, $C\in\mathbb{S}^n$

Table 2.1: Three types of linear conic optimization problems

We refer to the non-negative quantity $\langle x, z\rangle_{\mathbb{E}} \geq 0$ as the *duality gap*. When the duality gap is zero, so that inequality (2.1) holds as equality and $p^\star = d^\star$, we say that the primal and dual problems are in *strong duality*. From Proposition 2.1, it also follows that a pair of primal-dual feasible points $(x,z)\in\mathbb{E}\times\mathbb{E}$ realizing zero duality gap is necessarily a pair of primal-dual optimal solutions for the pair of problems (C_p, C_d).

This observation yields a set of first order optimality conditions, also known as the Karush-Kuhn-Tucker (KKT) conditions.

Definition 2.3 (KKT conditions). A primal-dual feasible point $(x,z) \in \mathbb{E} \times \mathbb{E}$ satisfies the *Karush-Kuhn-Tucker conditions* (KKT) for $(\mathrm{C}_p, \mathrm{C}_d)$ if

$$\begin{aligned} &\mathcal{A}(x) = b, \\ &\mathcal{A}^*(y) + z = c, \\ &(x,z) \in \mathcal{K} \times \mathcal{K}^*, \\ &\langle x, z\rangle_{\mathbb{E}} = 0, \end{aligned} \tag{KKT}$$

for some $y \in \mathbb{F}$.

If a primal-dual feasible point $(x,z) \in \mathbb{E} \times \mathbb{E}$ satisfies (KKT), then it is a primal-dual optimal solution for $(\mathrm{C}_p, \mathrm{C}_d)$. In general, for a pair of linear conic optimization problems only weak duality holds, so that (KKT) are only sufficient conditions for optimality. Under the condition of strict feasibility they are also necessary.

Definition 2.4 (Strict feasibility). We say that problem (C_p) is strictly feasible if there exists a point $x \in \mathbb{E}$ such that $\mathcal{A}(x) = b$ and $x \in \operatorname{relint} \mathcal{K}$. Analogously, we say that problem (C_d) is strictly feasible if there exists a point $z \in \mathbb{E}$ such that $\mathcal{A}^*(y) + z = c$ for some $y \in \mathbb{F}$ and $x \in \operatorname{relint} \mathcal{K}^*$.

Here, $\operatorname{relint} \mathcal{K}$ denotes the *relative interior* of $\mathcal{K}$:

$$\operatorname{relint} \mathcal{K} := \{x \in \mathcal{K} \mid \exists \epsilon > 0 \text{ s.t. } \mathcal{B}_x(\epsilon) \cap \operatorname{aff} \mathcal{K} \subseteq \mathcal{K}\},$$

where $\mathcal{B}_x(\epsilon)$ is a ball of radius ϵ centered on x and $\operatorname{aff} \mathcal{K}$ is the smallest affine space containing $\mathcal{K}$.

In the literature, strict feasibility is often refer to as Slater's condition, and it is regarded as a constraint qualification, a geometric condition of the feasible set that ensure that any local minimizer (which is global for convex problems) satisfies the first order KKT conditions for optimality.

Proposition 2.2 (Strict feasibility yields strong duality). *If problem (C_p) is strictly feasible then strong duality between (C_p) and (C_d) holds. Furthermore, the dual optimum (supremum) is attained. Conversely, if problem (C_d) is strictly feasible then strong duality between (C_p) and (C_d) holds. Furthermore, the primal optimum (infimum) is attained.*

For a proof see for example Section 5.3.2 in [34].

Scene 2 The fundamentals of SDP

Where we discover the many virtues of the young SDP twins. Unlike many stories of this kind, the relationship between the two brothers is normally very peaceful. Of course, like all siblings they sometimes have quarrels and disagreements, which weaken their dual brotherhood. Despite that, they usually get along very well, in particular after a good banquet, when the interiors of the two brothers are satisfactorily non-empty. In these occasions, their dual brotherhood is indeed strong...

In this section we review the standard properties of SDP. We recall that we write a semidefinite program in primal form as

$$\begin{aligned} p^\star = \inf_{X \in \mathbb{S}^n} \ & \langle C, X \rangle \\ \text{s.t.} \ & \mathcal{A}(X) = b, \\ & X \succeq 0, \end{aligned} \tag{P}$$

together with its dual, which we write in the form

$$\begin{aligned} d^\star = \sup_{y \in \mathbb{R}^m,\, Z \in \mathbb{S}^n} \ & \langle b, y \rangle \\ \text{s.t.} \ & \mathcal{A}^*(y) + Z = C, \\ & Z \succeq 0. \end{aligned} \tag{D}$$

The linear operator $\mathcal{A}$ maps a symmetric matrix $X \in \mathbb{S}^n$ to a vector in $\mathbb{R}^m$ defined by $(\langle A_1, X \rangle, \dots, \langle A_m, X \rangle)$, where $A_i \in \mathbb{S}^n$ are given matrices for $i \in [m]$ and $b \in \mathbb{R}^m$ is a vector, so that the variable matrix X must satisfy $\langle A_i, X \rangle = b_i$ for every $i \in [m]$. Finally, $\mathcal{A}^*$ is the linear operator adjoint to $\mathcal{A}$, and it is defined by setting $\mathcal{A}^*(y) = \sum_{i=1}^m A_i y_i$. Usually, one requires that the condition $\mathcal{A}(X) = b$ does not contain any redundant information.

Definition 2.5 (LICQ)**.** We say that the *linear independence constraint qualification* (LICQ) holds if the linear operator $\mathcal{A}$ is surjective. In other words, the m matrices $\{A_i\}_{i \in [m]}$ defining $\mathcal{A}$ are linearly independent in $\mathbb{S}^n$.

Without any loss of generality, we assume that LICQ holds.

Remark 1. Under LICQ, given a matrix $Z \in \mathbb{S}^n$ satisfying the dual constraint $\mathcal{A}^*(y) + Z = C$ for some $y \in \mathbb{R}^m$, y can be uniquely determined by solving the linear system $\mathcal{A}\mathcal{A}^*(y) = \mathcal{A}(C - Z)$, so that $y = (\mathcal{A}^*)^\dagger (C - Z)$, where the dagger denotes the pseudo-inverse. We exploit this fact and when discussing a dual solution (y, Z) we will often omit y and refer to a dual optimal solution simply as a matrix $Z \in \mathbb{S}^n$.

We call a matrix X satisfying the constraints of (P) a *primal feasible solution*, a matrix Z satisfying the constraints of (D) a *dual feasible solution*, a pair of matrices (X, Z) satisfying the constraints of (P, D) a *primal-dual feasible solution*. We call a matrix $X^\star$ such that $\langle C, X^\star \rangle = p^\star$ a *primal optimal solution*. Similarly, $Z^\star$ denotes a *dual optimal solution*, and we indicate with $(X^\star, Z^\star)$ a *primal-dual optimal solution*.

Clearly, the dual problem (D) is equivalent to

$$\begin{aligned} d^\star = \quad & \sup_{y \in \mathbb{R}^m} \ \langle b, y \rangle \\ & \text{s.t.} \quad C - \mathcal{A}^*(y) \succeq 0. \end{aligned} \tag{D$'$}$$

A relation of the type $C - \mathcal{A}^*(y) \succeq 0$ is called a Linear Matrix Inequality (LMI), so that an SDP can be thought as linear optimization with LMI constraints. The introduction of the matrix Z, often referred to as the *slack matrix* in the literature, allows to reformulate the dual problem symmetrically with respect to the primal problem.

The primal and dual problems are related by the Lagrangian function

$$\begin{aligned} \mathcal{L}(X, y) &= \langle C, X \rangle + \langle y, b - \mathcal{A}(X) \rangle \\ &= \langle b, y \rangle + \langle X, C - \mathcal{A}^*(y) \rangle, \end{aligned}$$

and can be derived via the standard techniques of Lagrangian duality in convex optimization. Observing that

$$\sup_{y \in \mathbb{R}^m} \mathcal{L}(X, y) = \begin{cases} \langle C, X \rangle & \text{if } \mathcal{A}(X) = b \\ +\infty & \text{otherwise} \end{cases},$$

and that similarly

$$\inf_{X \succeq 0} \mathcal{L}(X, y) = \begin{cases} \langle b, y \rangle & \text{if } C - \mathcal{A}^*(y) \succeq 0 \\ -\infty & \text{otherwise} \end{cases},$$

we can write

$$p^\star = \inf_{X \succeq 0} \left(\sup_{y \in \mathbb{R}^m} \mathcal{L}(X, y) \right) \tag{P}$$

and obtain the dual problem by formally exchange inf / sup:

$$d^\star = \sup_{y \in \mathbb{R}^m} \left(\inf_{X \succeq 0} \mathcal{L}(X, y) \right). \tag{D$'$}$$

This observation incidentally give us another proof of weak duality (2.1), which can now be derived from a direct application of the max-min inequality (e.g., Section 5.4 in [34]):

$$d^\star \le p^\star.$$

As discussed in the previous section, first order optimality conditions are available for any linear conic optimization problem, such that any pair of matrices satisfying these conditions is necessarily a pair of optimal solutions. These are the KKT conditions.

Definition 2.6 (KKT conditions for SDP)**.** A primal-dual pair of feasible solutions $(X,Z) \in \mathbb{S}^n \times \mathbb{S}^n$ satisfies the Karush-Kuhn-Tucker (KKT) conditions for (P, D) if, for some $y \in \mathbb{R}^m$,

$$\begin{aligned} &\mathcal{A}(X) = b \\ &\mathcal{A}^*(y) + Z = C \\ &X, Z \succeq 0 \\ &\langle X, Z\rangle = 0. \end{aligned} \tag{KKT}$$

Under primal-dual strict feasibility, the KKT conditions are also necessary.

Definition 2.7 (Strict feasibility for SDP)**.** We say that *strict feasibility* holds for (P) (or that (P) is *strictly feasible*) if there exists an interior point of the primal feasible region. That is, there exists a matrix $X \succ 0$ satisfying $\mathcal{A}(X) = b$. Similarly, strict feasibility holds for (D) (or (D) is *strictly feasible*) if there exists an interior point of the dual feasible region. That is, there exist $y \in \mathbb{R}^m$ and a matrix $Z \succ 0$ satisfying $\mathcal{A}^*(y) + Z = C$. If this property holds for both problems, we say that *primal-dual strict feasibility* holds.

Proposition 2.3 (Strict feasibility yields strong duality in SDP)**.** *If problem* (P) *is strictly feasible then strong duality between* (P) *and* (D) *holds, so that $p^\star = d^\star$. Furthermore, the dual optimum (supremum) is attained. Conversely, if problem* (D) *is strictly feasible then strong duality between* (P) *and* (D) *holds and the primal optimum (infimum) is attained.*

In other words, under strict feasibility solving (KKT) is equivalent to solving (P, D). Strong duality, and hence strict feasibility, has the important consequence that any pair of primal-dual optimal solutions is simultaneously diagonalizable.

Let us recall that for any pair of symmetric matrices (X,Z) with

$$\begin{aligned} \operatorname{rank} X &= r, \\ \operatorname{rank} Z &= s, \end{aligned}$$

can be diagonalized as follows:

$$\begin{aligned} X &= Q_X \operatorname{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0) Q_X^T, \\ Z &= Q_Z \operatorname{diag}(0, \ldots, 0, \sigma_{n-s+1}, \ldots, \sigma_n) Q_Z^T, \end{aligned} \tag{2.2}$$

where $\{\lambda_i\}_{i\in[n]}$ are the eigenvalues of X in decreasing order, so that $\lambda_1, \ldots, \lambda_r$ are the non-zero eigenvalues of X, and $\{\sigma_i\}_{i\in[n]}$ are the eigenvalues of Z in increasing order, so that $\sigma_{n-s+1}, \ldots, \sigma_n$ are the non-zero eigenvalues of Z.

It is easy to see that if X and Z are positive semidefinite, $\langle X, Z\rangle = 0$ is equivalent to $XZ = 0 = ZX$. In particular, X and Z commute. They are hence simultaneously diagonalizable (see e.g. Theorem 4.5.15 in [23]). That is, they share a basis of eigenvectors $\{q_i\}_{i\in[n]}$ and the orthogonal matrix $Q := [q_1, \ldots, q_n]$ is such that

$$Q = Q_X = Q_Z.$$

From $XZ = 0 = ZX$, we also deduce that if the primal and dual problems (P, D) are strictly feasible, any pair of primal-dual optimal solutions (X, Z) is such that the following inclusions are satisfied:

$$\begin{aligned} \operatorname{im} X &\subseteq \ker Z, \\ \operatorname{im} Z &\subseteq \ker X. \end{aligned}$$

Both inclusions imply that

$$\operatorname{rank} X + \operatorname{rank} Z \leq n,$$

which we refer to as the *complementarity* between the optimal solutions X and Z. Indeed, complementarity can be expressed by

$$\lambda_i \sigma_i = 0 \quad \text{for all } i \in [n],$$

in analogy with the complementarity slackness condition arising in LP between an optimal primal vector x and an optimal slack vector z (see row 1. of Table 2.1):

$$x_i z_i = 0 \quad \text{for all } i \in [n].$$

Definition 2.8 (Strict complementarity)**.** A primal-dual optimal point (X, Z) is said to be *strictly complementary* if

$$\operatorname{im} X = \ker Z \tag{2.3}$$

(or, equivalently, $\operatorname{im} Z = \ker X$).

In other terms,

$$\operatorname{rank} X + \operatorname{rank} Z = n,$$

and for every $i \in [n]$ exactly one of the two conditions $\lambda_i = 0$ and $\sigma_i = 0$ holds. We say that a primal-dual pair of SDPs as (P, D) satisfies *strict complementarity* if there exists a strictly complementary primal-dual optimal point (X, Z).

Before proceeding to the definition of non-degeneracy in SDP, we introduce the tangent space to the manifold of fixed rank matrices. This will be used in the definition of primal and dual non-degeneracy, as well as in Chapter 4 for the definition of the horizontal space.

Let us first define the set

$$\mathcal{M}_r := \{X \in \mathbb{S}^n \mid \operatorname{rank} X = r\}.$$

This is the manifold of fixed rank-r symmetric matrices. Since the eigenvalues of a matrix X are continuous functions of X, it is clear that, for $r > 0$, the boundary of $\mathcal{M}_r$ can be written as the disjoint unions of manifolds:

$$\partial \mathcal{M}_r = \mathcal{M}_0 \cup \cdots \cup \mathcal{M}_{r-1}.$$

Let now

$$\mathcal{M}_r^+ := \mathbb{S}_+^n \cap \mathcal{M}_r = \{X \in \mathbb{S}^n \mid X \succeq 0 \text{ and } \operatorname{rank}(X) = r\}$$

be the manifold of fixed rank-r positive semidefinite matrices.

The boundary of $\mathbb{S}_+^n$ can be decomposed as the disjoint unions of positive semidefinite matrices of fixed rank

$$\partial \mathbb{S}_+^n = \mathcal{M}_0^+ \cup \cdots \cup \mathcal{M}_{n-1}^+,$$

so that its interior coincides with $\mathcal{M}_n^+$. The tangent space of $\mathcal{M}_r$ at a point X is

$$\mathcal{T}_X = \left\{ Q_X \begin{bmatrix} U & V \\ V^T & 0 \end{bmatrix} Q_X^T \;\middle|\; U \in \mathbb{S}^r,\ V \in \mathbb{R}^{r\times(n-r)} \right\},$$

where Q_X is given by (2.2) and $r = \operatorname{rank} X$.

We are now ready to introduce the definitions of primal and dual non-degeneracy. All the definitions and results exposed below until the end of the section are due to [38].

Definition 2.9 (Non-degeneracy)**.** A primal feasible point X is *primal non-degenerate* if

$$\ker \mathcal{A} + \mathcal{T}_X = \mathbb{S}^n. \tag{2.4}$$

Symmetrically, a dual feasible point Z is *dual non-degenerate* if

$$\operatorname{im} \mathcal{A}^* + \mathcal{T}_Z = \mathbb{S}^n. \tag{2.5}$$

where

$$\mathcal{T}_Z = \left\{ Q_Z \begin{bmatrix} 0 & V \\ V^T & W \end{bmatrix} Q_Z^T \;\middle|\; W \in \mathbb{S}^s,\ V \in \mathbb{R}^{(n-s)\times s} \right\}$$

is the tangent space at Z in $\mathcal{M}_s$, where $s = \operatorname{rank}(Z)$ and Q_Z is given by (2.2). We say that a primal-dual feasible point (X, Z) is *non-degenerate* if X is primal non-degenerate and Z is dual non-degenerate.

Our interest in non-degeneracy is motivated by the following result:

Proposition 2.4.

1. *If $(X^\star, Z^\star)$ is a primal-dual non-degenerate optimal point then $(X^\star, Z^\star)$ is the unique primal-dual optimal point for (P, D).*

2. *Under strict complementarity, if $(X^\star, Z^\star)$ is a primal-dual unique optimal point then $(X^\star, Z^\star)$ is a non-degenerate primal-dual optimal point for (P, D).*

Remark 2. For a given point (X, Z), there exist linear algebraic conditions to check whether it is non-degenerate or not (see Theorems 6 and 9 in [38]).

Scene 3 Set-valued analysis for TV-SDP

Where we see our SDP heroes training hard in order to fullfill their dream and become a pair of TV-SDPs. They are now enrolled at the Royal Conic Institute of Cartography, where they put continuous effort in their studies under the experienced guidance of their french-polish professor, Dr. Painlevé-Kuratowski, a renowned map expert. After obtaining the Licence for Intrepid Cartographers of the Queen (LICQ) with full degree, they are ready for new challenges and exciting adventures...

We are interested in studying the trajectories of solutions to the primal TV-SDP

$$\begin{aligned} p_t^\star = \inf_{X\in\mathbb{S}^n} \ & \langle C_t, X\rangle \\ \text{s.t.} \ & \mathcal{A}_t(X) = b_t, \\ & X \succeq 0, \end{aligned} \tag{P$_t$}$$

with a time parameter $t \in (0,\tau)$ varying on a real interval.

For any given value of t the dual TV-SDP is

$$\begin{aligned} d_t^\star = \sup_{y\in\mathbb{R}^m, Z\in\mathbb{S}^n} \ & \langle b_t, y\rangle \\ \text{s.t.} \ & \mathcal{A}_t^*(y) + Z = C_t, \\ & Z \succeq 0. \end{aligned} \tag{D$_t$}$$

Definition 2.10 (Set-valued maps)**.** A *set-valued map* F from a set T to a set $\mathcal{X}$ maps a point $t \in T$ to a non-empty subset of $F(t) \subseteq \mathcal{X}$. In symbols:

$$\begin{aligned} F : T &\rightrightarrows \mathcal{X} \\ t &\mapsto F(t) \subseteq \mathcal{X}. \end{aligned}$$

We say that F is *single-valued* at $t \in T$ if $F(t)$ is a singleton. We say that F is *multi-valued* at $t \in T$ whenever $F(t)$ is neither empty nor a singleton.

Given the primal-dual pair of TV-SDPs (P_t, D_t), we define the primal and dual *feasible set-valued maps* for $t \in (0,\tau)$:

$$\begin{aligned} \mathcal{P}(t) &= \{X \in \mathbb{S}^n \mid \mathcal{A}_t(X) = b_t,\ X \succeq 0\}, \\ \mathcal{D}(t) &= \{Z \in \mathbb{S}^n \mid \mathcal{A}_t^*(y) + Z = C_t,\ y \in \mathbb{R}^m,\ Z \succeq 0\}, \end{aligned}$$

together with the primal and dual *optimal set-valued maps*:

$$\begin{aligned} \mathcal{P}^\star(t) &= \{X \in \mathcal{P}(t) \mid \langle C_t, X\rangle = p_t^\star\}, \\ \mathcal{D}^\star(t) &= \{Z \in \mathcal{D}(t) \mid \langle b_t, y\rangle = d_t^\star,\ \mathcal{A}_t^*(y) + Z = C_t,\ y \in \mathbb{R}^m\}. \end{aligned}$$

Continuity properties of set-valued maps can be defined in terms of outer and inner limits, leading to the notion of Painlevé–Kuratowski continuity. First, we introduce the notion of inner and outer limits of a set-valued map.

Definition 2.11 (Inner and outer limits)**.** Given a set-valued map $F : T \rightrightarrows \mathcal{X}$, its *inner limit* at $\hat{t} \in T$ is defined as

$$\liminf_{t \to \hat{t}} F(t) := \left\{ \hat{x} \;\middle|\; \forall \{t_k\}_{k=1}^{\infty} \subseteq T \text{ such that } t_k \to \hat{t},\ \exists \{x_k\}_{k=1}^{\infty} \subseteq \mathcal{X},\ x_k \to \hat{x} \text{ and } x_k \in F(t_k) \right\},$$

while its *outer limit* at $\hat{t} \in T$ is defined as

$$\limsup_{t \to \hat{t}} F(t) := \left\{ \hat{x} \;\middle|\; \exists \{t_k\}_{k=1}^{\infty} \subseteq T \text{ such that } t_k \to \hat{t},\ \exists \{x_k\}_{k=1}^{\infty} \subseteq \mathcal{X},\ x_k \to \hat{x} \text{ and } x_k \in F(t_k) \right\}.$$

Definition 2.12 (Painlevé–Kuratowski continuity)**.** Let $F : T \rightrightarrows \mathcal{X}$ be a set-valued map. We say that F is *outer semi-continuous* at $\hat{t} \in T$ if

$$\limsup_{t \to \hat{t}} F(t) \subseteq F(\hat{t}).$$

We say that F is *inner semi-continuous* at $\hat{t} \in T$ if

$$\liminf_{t \to \hat{t}} F(t) \supseteq F(\hat{t}).$$

Finally, we say that F is *Painlevé–Kuratowski continuous* at $\hat{t}$ if it is both outer and inner semi-continuous at $\hat{t}$.

Remark 3 (Continuity)**.** Note that a single-valued map $F : T \to \mathcal{X}$ is continuous in the usual sense at a point $x \in \mathcal{X}$ if and only if it is Painlevé–Kuratowski continuous at x as a multi-valued map $F : T \rightrightarrows \mathcal{X}$. Thus, without ambiguity, we will refer to Painlevé–Kuratowski continuity simply as continuity.

In the following, we list some continuity results on the feasible and optimal set-valued maps with $T = (0, \tau)$, for some $\tau > 0$. The proof of Theorem 2.2 in the primal version is an original contribution.

Theorem 2.1 (Example 5.8 in [75])**.** *If data $\mathcal{A}_t$, b_t, C_t are continuous functions of t, then the feasible set-valued maps $\mathcal{P}(t)$ and $\mathcal{D}(t)$ are outer semi-continuous at any $t \in T$.*

Theorem 2.2 (Theorem 2.12 in [61])**.** *Assume that the primal-dual pair of TV-SDPs (P_t, D_t) is strictly feasible at any $t \in T$, that the linear operator $\mathcal{A}_t$ satisfies LICQ for every $t \in T$, that the norm of $\mathcal{A}_t$ and the norm of its pseudo-inverse $\mathcal{A}_t^*(\mathcal{A}_t\mathcal{A}_t^*)^{-1}$ are uniformly bounded in t and that data $\mathcal{A}_t$, b_t, C_t are continuous functions of t. Then the set-valued maps $\mathcal{P}(t)$ and $\mathcal{D}(t)$ are inner semi-continuous for every $t \in T$.*

Proof. For the dual case, we refer to Lemma 1 in [58] for a version of this theorem where only the matrix C depends on the parameter and this dependence is linear. We prove the primal case in the more general case where the left hand side $\mathcal{A}_t$ is time-dependent and continuous and the right hand side b_t is continuous. The dual case can be proven in an analogous way. Fix $\hat{t} \in T$ and $\hat{X} \in \mathcal{P}(\hat{t})$. Given a sequence of times $\{t_k\}_{k=1}^{\infty}$ with $t_k \to \hat{t}$, we will construct a convergent sequence $X_k \to \hat{X}$ so that $X_k \in \mathcal{P}(t_k)$ for all sufficiently large values of k. If $\hat{X} \succ 0$ we define

$$X_k := \hat{X} + \mathcal{A}_{t_k}^*\left(\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*\right)^{-1}\left(b_{t_k} - b_{\hat{t}}\right).$$

The definition is well posed because under the assumptions of the theorem the operator $\mathcal{A}_{t_k}$ has full rank, thus $\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*$ is invertible. Clearly, $\mathcal{A}_{t_k}(X_k) = b_{t_k}$. Furthermore, we have that $\|X_k - \hat{X}\|_F = \|\mathcal{A}_{t_k}^*\left(\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*\right)^{-1}\left(b_{t_k} - b_{\hat{t}}\right)\|_F \leq C_{\mathcal{A}}\|b_{t_k} - b_{\hat{t}}\| \to 0$ for some constant $C_{\mathcal{A}}$ (which exists by the hypothesis of uniform boundedness) and by continuity of b_t, so that $X_k \to \hat{X}$ and $X_k \succeq 0$ for sufficiently large k. If $\hat{X} \succeq 0$ and its smallest eigenvalue $\lambda_{\min}(\hat{X})$ is zero, we define

$$X_k := (1-\alpha_k)\hat{X} + \alpha_k\bar{X} + \mathcal{A}_{t_k}^*\left(\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*\right)^{-1}\left(b_{t_k} - b_{\hat{t}}\right)$$

for a fixed $\bar{X} \in \mathcal{P}(\hat{t})$ such that $\bar{X} \succ 0$, which exists by the strict feasibility assumption, and for a sequence $\{\alpha_k\}_{k=1}^{\infty} \subseteq [0,1]$ which we will conveniently define later in the proof. Clearly, $\mathcal{A}_{t_k}(X_k) = b_{t_k}$ and hence we only need to prove that $X_k \succeq 0$ or, equivalently, that

$$\lambda_{\min}\left((1-\alpha_k)\hat{X} + \alpha_k\bar{X} + \mathcal{A}_{t_k}^*\left(\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*\right)^{-1}\left(b_{t_k} - b_{\hat{t}}\right)\right) \geq 0,$$

which, thanks to Weyl's inequality (see e.g Theorem 1 in [76], Section 6.7) holds if

$$\alpha_k\lambda_{\min}(\bar{X}) + \lambda_{\min}\left(\mathcal{A}_{t_k}^*\left(\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*\right)^{-1}\left(b_{t_k} - b_{\hat{t}}\right)\right) \geq 0.$$

Rearranging:

$$\alpha_k \geq -\frac{\lambda_{\min}\left(\mathcal{A}_{t_k}^*\left(\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*\right)^{-1}\left(b_{t_k} - b_{\hat{t}}\right)\right)}{\lambda_{\min}(\bar{X})}.$$

We then define $\alpha_k := \max\{0, \beta_k\}$, where

$$\beta_k := -\frac{\lambda_{\min}\left(\mathcal{A}_{t_k}^*\left(\mathcal{A}_{t_k}\mathcal{A}_{t_k}^*\right)^{-1}\left(b_{t_k} - b_{\hat{t}}\right)\right)}{\lambda_{\min}(\bar{X})}.$$

For sufficiently large k, $\beta_k \leq 1$, so that $\{\alpha_k\}_{k=1}^{\infty} \subseteq [0,1]$ and thus $X_k \in \mathcal{P}(t_k)$, since $\beta_k \to 0$, $\alpha_k \to 0$ and $X_k \to \hat{X}$. □

Theorems 2.1 and 2.2 show that the primal and dual feasible set-valued maps $\mathcal{P}(t)$ and $\mathcal{D}(t)$ are always continuous, under the assumptions of Theorem 2.2. We now investigate the inner and outer semi-continuity of the optimal set-valued maps. We have:

Theorem 2.3 (Theorem 8 in [77]). *If data $\mathcal{A}_t$, b_t, C_t are continuous functions of t and the primal and dual feasible set-valued maps are continuous, then the optimal set-valued maps $\mathcal{P}^\star(t)$ and $\mathcal{D}^\star(t)$ are outer semi-continuous at any $t \in T$.*

However, in general, it is not true that the optimal set-valued maps $\mathcal{P}^\star(t)$ and $\mathcal{D}^\star(t)$ are inner semi-continuous. Still, the set of $t \in T$ such that $\mathcal{P}^\star(t)$ or $\mathcal{D}^\star(t)$ fails to be inner semi-continuous, is of *first category*, i.e., countable and nowhere dense.

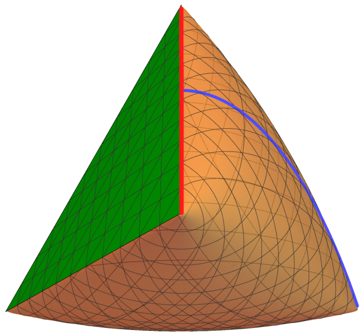
Theorem 2.4 (Theorem 5.55 in [75]). *The subset of points $t \in T$ at which $\mathcal{P}^\star(t)$ or $\mathcal{D}^\star(t)$ fails to be inner semi-continuous (and hence continuous) is the union of countably many sets that are nowhere dense in T. In particular, it has empty interior.*

Furthermore, if the optimal set is single-valued, then it is continuous everywhere. In order to show this, we first need to introduce a lemma which guarantees the local uniform boundedness of $\mathcal{P}^\star$ and $\mathcal{D}^\star$.

Lemma 2.1 (Lemma 3.2 in [4]). *If strict feasibility holds at any $t \in T$ and $\mathcal{A}_t$, b_t, C_t are continuous functions of t, then $\mathcal{P}^\star(t)$ and $\mathcal{D}^\star(t)$ are locally uniformly bounded at any $t \in T$.*

Proof. Since we assume that primal-dual strict feasibility holds at any $t \in T$, the assumptions of both Lemmas 3.1 and 3.2 in [4] are satisfied at any $t \in T$. □

Proposition 2.5 (Corollary 8.1 in [77]). *Assume that strict feasibility holds at any $t \in T$, so that by Lemma 2.1 $\mathcal{P}^\star(t)$ and $\mathcal{D}^\star(t)$ are locally uniformly bounded at any $t \in T$, and that $\mathcal{A}_t$, b_t, C_t are continuous functions of t. If $\mathcal{P}^\star(t)$ is single-valued at some $\hat{t}$, then $\mathcal{P}^\star(t)$ is continuous at $\hat{t}$. The same holds for $\mathcal{D}^\star(t)$.*



Chapter 3

Act II A complete classification of TV-SDP solutions

"Voi sapete quel che fa..."

Leporello in Don Giovanni (Act I, Scene 5)

The main goal of this chapter is to understand the properties of the trajectories of solutions to TV-SDPs, which would make it possible to design algorithms for TV-SDP with guarantees on their performance, as we will see in Chapter 4. Specifically, we want to characterize points of the trajectory of solutions to TV-SDP according to the local behavior of the trajectory at the point. This characterization was inspired by the textbook *Parametric optimization: singularities, pathfollowing and jumps* [63], where a classification of solutions to univariate parametric nonlinear constrained optimization problems is proposed. This chapter follows the paper *Parametric semidefinite programming: geometry of the trajectory of solutions* [61] resulted from the research pursued in the first part of this PhD project.

After investigating the regularity properties of the trajectory of solutions, we define six different types of points, according to the local behavior of the trajectory of the solutions at that point. Under very general and quite standard assumptions, we present a classification theorem, which states that only the types of points that we defined can appear. This provide the first complete classification of types of behavior of points making up the trajectory of solutions to TV-SDP. Furthermore, under some technical assumptions (the existence of a "non-singular time", see Definition 3.2), we will show that only three of the six types of points that we define can actually appear. Table 3.1 lists the types of points that we will define precisely later on. The benefits of this classification result

Problem assumptions	**Type of points**
TV-SDP with LICQ, continuous data, strict feasibility, and a *non-singular* time	Regular points Non-differentiable points Discontinuous isolated multiple points
TV-SDP with LICQ, continuous data, strict feasibility, without a *non-singular* time	Regular points Non-differentiable points Discontinuous isolated multiple points Discontinuous non-isolated multiple points Continuous bifurcation points Irregular accumulation points

Table 3.1: Assumptions on TV-SDP and associated possible type of points

are twofold: on the one hand, it increases the knowledge available in an interesting and relatively novel field of research; on the other hand, it sets the theoretical foundations for the actual development of algorithms to solve TV-SDP. For example, as we will see in the next chapter, in the path-following algorithm that we propose there, local information on the current iterate will be used, and one needs to know whether this information is reliable or not, i.e., whether the solution is expected to behave regularly or not.

Scene 1 Regularity properties of TV-SDP

Where our heroes, the SDP prince and its brother the dual prince, start moving their first steps out of the native kingdom. By now, the two protagonists have turned into a couple of mature and yet adventurous TV-SDPs, ready to intrepidly explore the enchanted realm of Parametric Optimization. Thanks to the uniqueness of their virtues and the strict complementarity of their dual brotherhood, the beginning of our heroes' journey is indeed a smooth one...

Our purpose is to study the behavior of the trajectory of the solutions to TV-SDPs. Around points of the trajectory satisfying strict complementarity and uniqueness, by means of the implicit function theorem, we will show that the trajectory defines a smooth curve (Theorem 3.1). When this fails to happen, a number of irregular behaviors may arise. The main result of this chapter (Theorem 3.3) consists of a complete classification of such points. So far, to the best of our knowledge, a complete classification of types of behavior of points making up the trajectory of solutions has not been proposed. Here, we suggest one based on a purely logical construction, whose definitions use set-valued analysis. In particular, we use the Painlevé–Kuratowski extension of the notion of continuity to the case of set-valued functions, as exposed in the previous chapter, so as to reason about continuity properties at values of the parameter when there are multiple solutions.

A first contribution of this thesis is the definition of six types of points. Our approach was deeply inspired by [63, Chapter 2] where a classification of solutions to univariate parametric nonlinear constrained optimization problems is proposed. There, critical points satisfying KKT conditions are considered. Under precise algebraic conditions, these points are "non-degenerate" (see Remark 8). The local behavior of such points is then shown to be regular. If a critical point is instead "degenerate" then, according to which algebraic condition is satisfied, the point is classified into four different types. Our approach is the same in spirit, in that we also start by considering algebraic conditions ensuring a regular behavior. As a main difference, we classify irregular points according to the behavior of the trajectory of solutions at the point considered rather than according to different sets of algebraic conditions (see Remark 8).

Let us recall that we are considering a TV-SDP problem in its primal standard form

$$\begin{aligned} p_t^\star = \inf_{X\in\mathbb{S}^n} \ & \langle C_t, X\rangle \\ \text{s.t.}\ & \mathcal{A}_t(X) = b_t, \\ & X \succeq 0, \end{aligned} \tag{P$_t$}$$

along with its dual problem

$$\begin{aligned} d_t^\star = \sup_{y\in\mathbb{R}^m, Z\in\mathbb{S}^n} \ & \langle b_t, y\rangle \\ \text{s.t.}\ & \mathcal{A}_t^*(y) + Z = C_t, \\ & Z \succeq 0, \end{aligned} \tag{D$_t$}$$

with a time parameter $t \in (0,\tau)$ varying on a open and bounded real interval. We consider an open interval so that both left and right limits can be defined when arguing about continuity and differentiability.

In this section, we are going to show that the existence of a unique pair of strictly complementary primal and dual optimal solutions at a value of the time parameter $\hat{t} \in (0,\tau)$ implies that there is a neighborhood of $\hat{t}$ where both the primal and dual optimal trajectory have a regular behavior. Next, we observe that under fairly weak assumptions, among which the existence of a generic non-singular point in the parameterization interval, the number of points where strict complementarity or uniqueness is lost is finite. Given the primal-dual pair of TV-SDPs (P_t, D_t), we denote a primal-dual feasible solution by (X, Z, t), which we will often refer to as a *point* of the trajectory of solution. If at a fixed value of the parameter $\hat{t} \in (0,\tau)$ there exists a primal-dual non-degenerate optimal solution $(X^\star, Z^\star)$, then, by Proposition 2.4 $(X^\star, Z^\star)$ is a unique primal-dual optimal solution, and by Proposition 2.5, around $\hat{t}$ the primal and dual optimal set-valued maps are continuous single-valued functions. Under strict complementarity, these functions are analytic. In the following, we provide details of this fact.

We begin by adopting the following standard assumptions:

(A1) Data $\mathcal{A}_t$, b_t, C_t are continuous functions of t.

(A2) The linear operator $\mathcal{A}_t$ satisfies LICQ for any $t \in (0, \tau)$. Furthermore, $\mathcal{A}_t$ and its pseudo-inverse $(\mathcal{A}_t^*)^\dagger = \mathcal{A}_t^*\big(\mathcal{A}_t\mathcal{A}_t^*\big)^{-1}$ have a uniformly bounded norm for any $t \in (0, \tau)$.

(A3) Problem (P_t) and its dual (D_t) are strictly feasible for every $t \in (0, \tau)$.

Assumption A1 is quite general compared to those usually found in the TV-SDP literature, where the data are often assumed to vary linearly with respect to the time parameter. This linearity assumption is standard when one studies sensitivity properties, so that the perturbation can be assumed to be linear. Instead, our purpose is to give a geometric characterization of the points of the trajectory of solutions, in which case we can keep a high degree of generality by just assuming continuity of the data, without any further differentiability requirement.

Assumption A2 allows us to describe the dual solution just in terms of matrix Z (see Remark 1). The assumption of uniform boundedness is needed to ensure the inner semi-continuity of the feasible set-valued maps (see Theorem 2.2).

Assumption A3 is standard in the SDP literature [52, 51, 58]. Strict feasibility guarantees that the primal and dual optimal sets $\mathcal{P}^\star(t)$ and $\mathcal{D}^\star(t)$ are non-empty and bounded for any $t \in (0, \tau)$ (see Lemma 3.2 in [78]). Checking strict feasibility of a given SDP can be done by solving another SDP and checking whether its optimal value is positive or not (see for example [79], Theorem 3.1 and 3.5).

Summarizing, assumptions A1, A2, and A3 ensure that:

- There is no duality gap: $p_t^\star = d_t^\star$ for all $t \in (0, \tau)$.
- In (P_t) and (D_t) the infimum and the supremum are attained.
- The primal and dual optimal faces $\mathcal{P}^\star(t)$, $\mathcal{D}^\star(t)$ are non-empty and bounded for all $t \in (0, \tau)$. In other words, (P_t) and (D_t) are both feasible and bounded.
- The optimal set-valued maps are outer semi-continuous at any $t \in (0, \tau)$.

- The subset of $(0,\tau)$ where the optimal set-valued map fails to be inner semi-continuous has empty interior and it is the union of countably many sets that are nowhere dense in $(0,\tau)$.

The last two facts follow immediately from Theorems 2.3 and 2.4.

The optimality conditions (KKT) for (X,Z,t) to be an optimal solution of $(\mathrm{P}_t,\mathrm{D}_t)$ at a fixed value of the parameter $t\in(0,\tau)$ can be equivalently written as

$$F(X,y,Z,t) := \begin{bmatrix} \mathcal{A}_t\,\mathrm{svec}(X) - b_t, \\ \mathcal{A}_t^T y + \mathrm{svec}(Z) - \mathrm{svec}(C_t) \\ \frac{1}{2}\,\mathrm{svec}\,(XZ+ZX) \end{bmatrix} = 0, \tag{3.1}$$

$$X, Z \succeq 0 \tag{3.2}$$

for some $y\in\mathbb{R}^m$, where $\mathcal{A}_t := \left(\mathrm{svec}(A_{1,t}),\ldots,\mathrm{svec}(A_{m,t})\right)^T$ and $\mathrm{svec}(X)$ denotes a linear map stacking the upper triangular part of X, where the off-diagonal entries are multiplied by $\sqrt{2}$:

$$\mathrm{svec}(X) := \left(X_{11},\sqrt{2}X_{12},\ldots,\sqrt{2}X_{1n},X_{22},\sqrt{2}X_{23},\ldots,\sqrt{2}X_{2n},\ldots,X_{nn}\right)^T$$

so that $\langle X,X\rangle = \mathrm{svec}(X)^T\,\mathrm{svec}(X)$.

Theorem 7 in [80] shows that for a generic data tuple $(\mathcal{A},b,C)$ the number of solutions (X,y,Z) for (3.1) is fixed and finite, depending only on the dimensions of the problem.

Definition 3.1 (Singular solutions)**.** We say that a primal-dual solution (X,y,Z) is *singular* at a time $\hat{t}$ if the Jacobian with respect to (X,y,Z) of F at $(X,y,Z,\hat{t}\,)$

$$\mathcal{J}(X,y,Z,\hat{t}\,) = \begin{bmatrix} \mathcal{A}_{\hat{t}} & 0 & 0 \\ 0 & \mathcal{A}_{\hat{t}}^T & I_{\frac{n(n+1)}{2}} \\ Z\otimes_s I_n & 0 & I_n\otimes_s X \end{bmatrix} \tag{3.3}$$

is not invertible, where $\otimes_s$ denotes the symmetric Kronecker product between two $n\times n$ matrices M_1 and M_2 and is defined by

$$(M_1\otimes_s M_2)\,\mathrm{svec}(S) = \frac{1}{2}(M_1SM_2^T + M_2SM_1^T) \quad \text{for any } S\in\mathbb{S}^n.$$

Otherwise, we say that (X,y,Z) is *non-singular* at $\hat{t}$.

Definition 3.2 (Singular times)**.** We say that a time $\hat{t}$ is *singular* if there exists a singular point (X, y, Z) at $\hat{t}$ such that $F(X, y, Z, \hat{t}\,) = 0$. Otherwise, we say that $\hat{t}$ is *non-singular*. Furthermore, we say that a non-singular time $\hat{t}$ is *generic* if the data tuple $(\mathcal{A}_{\hat{t}}, b_{\hat{t}}, C_{\hat{t}})$ is generic, so that the number of solutions for (3.1) matches the generic number of solutions.

Note that if $\hat{t}$ is non-singular, every point (X, y, Z) such that $F(X, y, Z, \hat{t}\,) = 0$ is non-singular at $\hat{t}$.

The following lemma gives equivalent conditions for a primal-dual optimal point (X, Z) to be non-singular at a time $\hat{t}$. This result turns out to be a fundamental tool to reach our classification goal.

Lemma 3.1 (Theorem 3.1. in [81] and [82])**.** *A primal-dual optimal point (X, Z) is non-singular at a time $\hat{t}$ if and only if (X, Z) is a strictly complementary and non-degenerate primal-dual optimal solution for* $(\mathrm{P}_{\hat{t}}, \mathrm{D}_{\hat{t}})$.

Note that under strict complementarity, part 2 of Proposition 2.4 holds. Therefore, the Jacobian of F is non-singular at an optimal primal-dual solution (X, Z, t) if and only if (X, Z) is a *unique* primal-dual optimal solution satisfying strict complementarity. We use this result in the following theorem.

Theorem 3.1. *Let $(\mathrm{P}_t, \mathrm{D}_t)$ be a primal-dual pair of TV-SDPs parametrized over the time interval $(0, \tau)$ such that primal-dual strict feasibility holds for any $t \in (0, \tau)$ and assume that the data $\mathcal{A}_t, b_t, C_t$ are continuously differentiable functions of t. Let $\hat{t} \in (0, \tau)$ be a fixed value of the time parameter and suppose that $(X^\star, Z^\star)$ is a unique primal-dual optimal and strictly complementary solution for $(\mathrm{P}_{\hat{t}}, \mathrm{D}_{\hat{t}})$.*

Then there exists $\varepsilon > 0$ and a unique continuously differentiable mapping $(X^\star(\cdot), Z^\star(\cdot))$ defined on $(\hat{t} - \varepsilon, \hat{t} + \varepsilon)$ such that $(X^\star(t), Z^\star(t))$ is a unique and strictly complementary primal-dual optimal point to $(\mathrm{P}_t, \mathrm{D}_t)$ for all $t \in (\hat{t} - \varepsilon, \hat{t} + \varepsilon)$.

Proof. By primal-dual strict feasibility, for each $t \in (0, \tau)$ the pair of problems (P_t, D_t) must have at least a primal-dual feasible and optimal solution, which we denote as $(X_t^\star, y_t^\star, Z_t^\star)$, which necessarily solves the KKT system (3.1)-(3.2). In particular, we have $F(X_t^\star, y_t^\star, Z_t^\star, t) = 0$. By Lemma 3.1, the assumptions of strict complementarity and uniqueness ensure that at $\hat{t}$ we can apply the Implicit Function Theorem (see, e.g., Theorem 3.3.1 in [83]), so that there exists $\varepsilon' > 0$ and a continuously differentiable curve $(X(\cdot), y(\cdot), Z(\cdot))$ on $(\hat{t} - \varepsilon', \hat{t} + \varepsilon')$ such that $(X(t), y(t), Z(t), t)$ is the unique solution of $F(X(t), y(t), Z(t), t) = 0$ for all $t \in (\hat{t} - \varepsilon', \hat{t} + \varepsilon')$. Hence for $t \in (\hat{t} - \varepsilon', \hat{t} + \varepsilon')$, $(X(t), y(t), Z(t), t)$ must coincide with $(X_t^\star, y_t^\star, Z_t^\star, t)$, since the latter both solves (3.1) and

(3.2), and it is therefore a feasible and unique primal-dual optimal solution for ($\mathrm{P}_t, \mathrm{D}_t$). For consistency of notation, we can now write $(X(t), y(t), Z(t), t) \equiv (X^\star(t), y^\star(t), Z^\star(t), t)$. It now remains to show the strict complementarity of $(X^\star(t), Z^\star(t))$. Due to the assumed strict complementarity of $(X^\star(\hat{t}), Z^\star(\hat{t}))$, we have $\lambda_i(X^\star(\hat{t})) \cdot \lambda_i(Z^\star(\hat{t})) = 0$ and $\lambda_i(X^\star(\hat{t})) + \lambda_i(Z^\star(\hat{t})) > 0$, where $\lambda_i(\cdot)$ denotes the i-th smallest eigenvalue of a matrix. By continuity of the eigenvalues, for ε small enough, nonzero eigenvalues remains nonzero for $t \in (\hat{t} - \varepsilon, \hat{t} + \varepsilon)$, and the perturbed solutions remain strictly complementary. $\square$

By adding more assumptions, one can further improve the information given by Theorem 2.4 on the cardinality of the singular points set and demonstrate that the number of singular points of (3.1) is finite.

Theorem 3.2 (Proposition 5 in [58]). *Let ($\mathrm{P}_t, \mathrm{D}_t$) be a primal-dual pair of TV-SDPs parametrized over a time interval $(0, \tau)$ such that data $\mathcal{A}_t, b_t, C_t$ are polynomial functions of t. Furthermore, assume that there exists a generic non-singular time. Then the set of values of the time parameter t at which the primal-dual optimal point is either not unique or not strictly complementary is finite.*

Proof. Let us elaborate on the proof given by [58] to prove their Proposition 5. We first define the set

$$\mathcal{F} := \{(X, y, Z, t) \in \mathbb{C}^{\frac{n(n+1)}{2}} \times \mathbb{C}^m \times \mathbb{C}^{\frac{n(n+1)}{2}} \times \mathbb{C} \mid F(X, y, Z, t) = 0,\ \det(J_F(X, y, Z, t)) = 0\},$$

which is a *constructible set* (see, e.g., Section 1.1 in [84]). The projection of a constructible set is a constructible set itself (Theorem 1.32 in [84]), so that the projection of $\mathcal{F}$ on the t coordinate

$$\mathcal{F}_{proj} = \left\{ t \in \mathbb{C} \mid \exists (X, y, Z, t) \in \mathcal{F} \right\}$$

is a constructible set in $\mathbb{C}$. At this point, we exploit the fact that any constructible set of $\mathbb{C}$ is either a finite set or the complement of a finite set (Exercise 1.3 in [84]). By the hypothesis that there exists a non-singular time $\hat{t}$ it follows that the complement of $\mathcal{F}_{proj}$ contains $\hat{t}$ and thus, from the implicit function theorem, $F(X, y, Z, t) = 0$ and $\det(J_F(X, y, Z, t)) \neq 0$ for all t in an open neighborhood of $\hat{t}$. By the genericity assumption, in this neighborhood the number of solutions to (3.1) is constant, so that indeed *all* the solution of $F(X, y, Z, t) = 0$ are non-singular. This neighborhood is contained in the complement of $\mathcal{F}_{proj}$ and it is not finite (it is a open interval with non-empty interior).

Hence $\mathcal{F}_{proj}$ is a finite set. Since

$$\{t \in (0,\tau) \mid \exists (X,y,Z) \in \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n \text{ s.t. } F(X,y,Z,t) = 0,\ \det(J_F(X,y,Z,t)) = 0\}$$

is contained in $\mathcal{F}_{proj}$, the set of values of the parameter at which there exists a singular point for the Jacobian of F is also finite. Application of Lemma 3.1 yields the final result. □

Thus, under the assumption of Theorem 3.2, the values of t at which strict complementarity or uniqueness of the primal-dual solution is lost is finite. In particular, the values of t at which $\mathcal{P}^\star(t)$ or $\mathcal{D}^\star(t)$ fails to be inner semi-continuous (and hence fails to be continuous) are finite. It also implies that wherever $\mathcal{P}^\star(t)$ defines a continuous curve of unique optima, the values of t at which $\mathcal{P}^\star(t)$ fails to be differentiable are finite. The same holds for $\mathcal{D}^\star(t)$.

Scene 2 Six types of points in the trajectory of solutions

Where we witness the first misadventure of the TV-SDP princes. After a long but safe journey, following a rather narrow and plain path, they enter the Singular Wood, a big enchanted forest in the realm of Parametric Optimization. There, they meet the Degenerate Witch, disguised as a charming and lovely maiden, who seduces the two princes. Madly in love, they start a dreadful dispute, contending for the favors of the mysterious lady. This seriously damages their unity and strict complementarity. When the fight is over, they realize with dismay that they lost the way...

Equipped with the results of the previous section, we introduce a classification into six different types of optimal solutions according to the behavior of the optimal set-valued map at these points. Our purpose is to study irregularities arising after an interval where the optimal set-valued map has a regular behavior. We hence classify points for which the optimal set-valued map on a left neighborhood is unique and thus continuous.

Consequently, the focus of our study is put on values $\hat{t}$ of an open time parametrization interval $(0,\tau)$ at which strict complementarity or uniqueness of the primal-dual optimal point is lost. Under the assumptions of Theorem 3.2, such points are finite. There, the trajectory described by the primal and dual optimal sets can exhibit a restricted number of irregular behaviors, by which we mean any situation that differs from the solution following a uniquely well-defined smooth curve. Describing these situations is the goal of this section. If, instead, Theorem 3.2 does not hold, the number of possible types of irregular behaviors grows. In our main Theorem 3.3, we provide a complete classification of these behaviors under both cases. The object of our study is the trajectory of solutions to the primal TV-SDP (P_t), that is, the primal optimal set-valued map $\mathcal{P}^\star$. Both results that we present in this chapter, Theorem 3.3 and Theorem 3.4, can be clearly transposed to the dual case.

We start by giving the definitions of the six types of points in the trajectory of solutions. Let (P_t, D_t) be a primal-dual pair of TV-SDPs parametrized over a time interval $(0,\tau)$. For a fixed $\hat{t} \in (0,\tau)$, we consider a primal optimal point $(X^\star, \hat{t})$ for ($\mathrm{P}_{\hat{t}}$). Based on the behavior of the primal optimal set-valued map $\mathcal{P}^\star(t)$ at $\hat{t}$, we distinguish between six cases. This can be done analogously for the dual case.

Definition 3.3 (Regular point)**.** We say that $(X^\star, \hat{t})$ is a *regular point* if $\mathcal{P}^\star(\hat{t}) = \{X^\star\}$ and there exists $\varepsilon > 0$ such that

- $\mathcal{P}^\star(t)$ is single-valued and continuous for every $t \in (\hat{t}-\varepsilon, \hat{t}+\varepsilon)$, for some $\varepsilon > 0$,
- $\mathcal{P}^\star(t)$ is differentiable at $\hat{t}$.

Remark 4. Note that a primal optimal point $(X^\star, \hat{t})$ for ($\mathrm{P}_{\hat{t}}$) for which there exists a dual optimal point $(Z^\star, \hat{t})$ for $(D_{\hat{t}})$ such that $(X^\star, Z^\star, \hat{t})$ is a non-singular point for ($\mathrm{P}_{\hat{t}}$, $\mathrm{D}_{\hat{t}}$), is necessarily a regular point. This follows directly from Theorem 3.1 and Lemma 3.1. The converse does not hold in general.

Definition 3.4 (Non-differentiable point)**.** We say that $(X^\star, \hat{t})$ is a *non-differentiable point* if $\mathcal{P}^\star(\hat{t}) = \{X^\star\}$ and there exists $\varepsilon > 0$ such that

- $\mathcal{P}^\star(t)$ is single-valued and continuous for every $t \in (\hat{t}-\varepsilon, \hat{t}+\varepsilon)$,
- $\mathcal{P}^\star(t)$ is *not* differentiable at $\hat{t}$.

Definition 3.5 (Discontinuous isolated multiple point)**.** We say that $(X^\star, \hat{t})$ is a *discontinuous isolated multiple point* if $X^\star \in \mathcal{P}^\star(\hat{t})$ and there exists $\varepsilon > 0$ such that

- $\mathcal{P}^\star(t)$ is single-valued and continuous for every $t \in (\hat{t}-\varepsilon, \hat{t}) \cup (\hat{t}, \hat{t}+\varepsilon)$,
- $\mathcal{P}^\star(t)$ is multi-valued at $\hat{t}$.

Definition 3.6 (Discontinuous non-isolated multiple point)**.** We say that $(X^\star, \hat{t})$ is a *discontinuous non-isolated multiple point* if $X^\star \in \mathcal{P}^\star(\hat{t})$ and there exists $\varepsilon > 0$ such that

- $\mathcal{P}^\star(t)$ is continuous at any $t \in (\hat{t} - \varepsilon, \hat{t}) \cup (\hat{t}, \hat{t} + \varepsilon)$,
- $\mathcal{P}^\star(t)$ is single-valued for every $t \in (\hat{t} - \varepsilon, \hat{t})$,
- $\mathcal{P}^\star(t)$ is multi-valued for every $t \in [\hat{t}, \hat{t} + \varepsilon)$.

Remark 5. Let $(X_1^\star, \hat{t}_1)$ be a discontinuous isolated multiple point and $(X_2^\star, \hat{t}_2)$ a discontinuous non-isolated multiple point. Then by definition the optimal solution is not unique neither at $\hat{t}_1$ nor at $\hat{t}_2$. Thus, a loss of inner semi-continuity of the optimal set-valued map $\mathcal{P}^\star(t)$ must occur both at $\hat{t}_1$ and at $\hat{t}_2$. However, while for any $\varepsilon > 0$ the set of points $t \in (\hat{t}_2 - \varepsilon, \hat{t}_2 + \varepsilon)$ where the optimal set $\mathcal{P}^\star(t)$ is multi-valued has a non-empty interior, there always exists a $\bar{\varepsilon} > 0$ such that the set of points $t \in (\hat{t}_1 - \bar{\varepsilon}, \hat{t}_1 + \bar{\varepsilon})$ where the optimal set $\mathcal{P}^\star(t)$ is multi-valued has empty interior. This observation suggests the choice of the terms “isolated" and “non-isolated".

Definition 3.7 (Continuous bifurcation point)**.** We say that $(X^\star, \hat{t})$ is a *continuous bifurcation point* if $\mathcal{P}^\star(\hat{t}) = \{X^\star\}$ and there exists $\varepsilon > 0$ such that

- $\mathcal{P}^\star(t)$ is continuous at any $t \in (\hat{t} - \varepsilon, \hat{t} + \varepsilon)$,
- $\mathcal{P}^\star(t)$ is single-valued for every $t \in (\hat{t} - \varepsilon, \hat{t}]$,
- $\mathcal{P}^\star(t)$ is multi-valued for every $t \in (\hat{t}, \hat{t} + \varepsilon)$.

In particular, there exist at least two distinct continuous curves

$$\begin{aligned} X_1 : (\hat{t}, \hat{t} + \varepsilon) &\to \mathbb{S}^n \\ t &\mapsto X_1(t) \end{aligned} \qquad \begin{aligned} X_2 : (\hat{t}, \hat{t} + \varepsilon) &\to \mathbb{S}^n \\ t &\mapsto X_2(t) \end{aligned}$$

such that $X_1(t)$ and $X_2(t)$ are two distinct points of $\mathcal{P}^\star(t)$ for every $t \in (\hat{t}, \hat{t} + \varepsilon)$ and $\lim_{t \to \hat{t}^+} X_1(t) = \lim_{t \to \hat{t}^+} X_2(t) = X^\star$. In this sense, a continuous bifurcation point can be thought as a continuous loss of uniqueness from a single branch into two or more branches.

Definition 3.8 (Irregular accumulation point)**.** We say that $(X^\star, \hat{t})$ is an *irregular accumulation point* if $X^\star \in \mathcal{P}^\star(\hat{t})$ and there exists $\varepsilon > 0$ such that

- $\mathcal{P}^\star(t)$ is single-valued and continuous for every $t \in (\hat{t} - \varepsilon, \hat{t})$

and for any $\delta > 0$ at least one of the following is true:

- there exists a sequence of times $\{t_k\}_{k=1}^{\infty} \subseteq (\hat{t}, \hat{t} + \delta)$ at which a loss of inner semi-continuity occurs and $\lim_{k\to\infty} t_k = \hat{t}$. At these times, either a discontinuous isolated multiple point or a discontinuous non-isolated multiple point appears.

- there exists a sequence of times $\{t_k\}_{k=1}^{\infty} \subseteq (\hat{t}, \hat{t}+\delta)$ at which a continuous bifurcation occurs and $\lim_{k\to\infty} t_k = \hat{t}$.

When convenient, instead of saying that $(X^\star, \hat{t}\,)$ is a regular point, we will say that $X^\star$ is a regular point at $\hat{t}$. The same applies to all the other types of points that we defined.

Remark 6. The above definitions consider points whose sufficiently small *left* time neighborhood consists of all regular points. By a change of sign of the parameter, the definition clearly extends to points whose sufficiently small *right* time neighborhood consists of all regular points.

Remark 7 (Existence of a continuous selection)**.** The optimal set-valued map is continuous in a neighborhood of a regular, non-differentiable, or a continuous bifurcation point. Instead, at a discontinuous isolated or non-isolated multiple point (Definitions 3.5 and 3.6), a loss of inner semi-continuity occurs. For such points $(X^\star, \hat{t}\,)$ it holds $\liminf_{t\to\hat{t}^-} \mathcal{P}^\star(t) \neq \mathcal{P}^\star(\hat{t}\,)$. However, in both cases, clearly only one of the following is true:

$$\begin{aligned} &(A) \quad \lim_{t\to\hat{t}^+} \mathcal{P}^\star(t) = \mathcal{P}^\star(\hat{t}\,), \\ &(B) \quad \liminf_{t\to\hat{t}^+} \mathcal{P}^\star(t) \neq \mathcal{P}^\star(\hat{t}\,). \end{aligned}$$

In case (A), one can select a continuous curve $(\hat{t} - \varepsilon, \hat{t} + \varepsilon) \ni t \mapsto X(t) \in \mathbb{S}^n$ such that $X(t) \in \mathcal{P}^\star(t)$ for every $t \in (\hat{t} - \varepsilon, \hat{t} + \varepsilon)$, while in case (B) such a curve does not exist. Furthermore, for a discontinuous isolated multiple point under case (A), such a curve is unique. Also note that in case (A) it might be impossible to select a curve that is differentiable at $\hat{t}$.

Remark 8 (Comparison with [63])**.** The definition of the six different types of points was inspired by [63, Chapter 2], where a classification of solutions to univariate parametric non-linear constrained optimization problems was proposed. There, critical primal-dual points satisfying first-order optimality (or KKT) conditions for a given parametric non-linear optimization problem are classified. These points are defined as *non-degenerate* if strict complementarity holds as well as the invertibility of the Hessian of the Lagrangian of the considered problem restricted to the tangent space at the point. We remark that this notion of non-degeneracy does not coincide with that of primal and dual non-degeneracy defined according to Definition 2.9. However, one can still identify an algebraic resemblance between primal non-degeneracy as we define it here and the non-singularity of the Hessian of the Lagrangian.

In the terminology that we used, the notion of *non-degeneracy* adopted by Jongen in [63] is analogous to *non-singularity*, as defined in Definition 3.1, as they both guarantee the applicability of the implicit function theorem, hence ensuring a regular behavior (Theorem 2.4.2 in [63]). Around these points the optimal set can be parametrized by means of a single parameter and the parameterization is a differentiable map. If a critical point is instead *degenerate* then, according to which algebraic condition is not satisfied by such points, these are classified in four different types. Instead, we classified irregular points according to the behavior of the trajectory of solutions at the point considered, focusing on the possible local topological structure of points

Theorem 3.3 (A complete classification)**.** *Let* $(\mathrm{P}_t, \mathrm{D}_t)$ *be a primal-dual pair of TV-SDPs parametrized over a time interval* $(0,\tau)$ *such that assumptions A1, A2, and A3 hold and consider a time* $\hat{t} \in (0,\tau)$ *and an optimal solution* $X^\star \in \mathcal{P}^\star(\hat{t}\,)$*. If* $\mathcal{P}^\star(t)$ *is unique for every* $t \in (\hat{t} - \varepsilon', \hat{t}\,)$ *for some* $\varepsilon' > 0$*, then* $(X^\star, \hat{t}\,)$ *must be a point of a type defined in Definitions 3.3, 3.4, 3.5, 3.6, 3.7, or 3.8. The same holds for* $\mathcal{D}^\star(t)$*.*

Proof. First, let $\hat{t} \in (0,\tau)$ and $X^\star \in \mathcal{P}^\star(\hat{t}\,)$. By hypothesis, there exists $\varepsilon' > 0$ such that $\mathcal{P}^\star(t)$ is single-valued and hence, by Proposition 2.5, continuous for every $t \in (\hat{t} - \varepsilon', \hat{t}\,)$. Let us perform a first binary case partition:

A $\mathcal{P}^\star(\hat{t}\,)$ is a single-valued (and thus equal to $\{X^\star\}$).

B $\mathcal{P}^\star(\hat{t}\,)$ is multi-valued.

Then, we also define a three-way case partition, independent from the previous one:

1 there exists $\varepsilon'' > 0$ such that $\mathcal{P}^\star(t)$ is single-valued for every $t \in (\hat{t}, \hat{t} + \varepsilon'')$.

2 there exists $\varepsilon'' > 0$ $\mathcal{P}^\star(t)$ is multi-valued for every $t \in (\hat{t}, \hat{t} + \varepsilon'')$.

3 for every $\delta > 0$ there exists $t', t'' \in (\hat{t}, \hat{t} + \delta)$ such that $\mathcal{P}^\star(t')$ is single-valued and $\mathcal{P}^\star(t'')$ is multi-valued.

Combining the two partitions, we obtain one consisting of six cases:

A1 in this case $\mathcal{P}^\star(t)$ is a single-valued function defined in $(\hat{t}-\varepsilon, \hat{t}+\varepsilon)$, where $\varepsilon := \min\{\varepsilon', \varepsilon''\}$, which is hence continuous by Proposition 2.5. According to whether $\mathcal{P}^\star(t)$ is differentiable at $\hat{t}$ or not, $(X^\star, \hat{t}\,)$ is a regular point or a non-differentiable point.

A2 if there exists $\varepsilon'' > 0$ such that $\mathcal{P}^\star(t)$ is continuous at any $t \in (\hat{t}-\varepsilon', \hat{t}+\varepsilon'')$ then by definition $(X^\star, \hat{t}\,)$ is a continuous bifurcation point (Definition 3.7). Otherwise, for every $k \in \mathbb{N}$ there must exist a point $t_k \in (\hat{t}, \hat{t}+\frac{1}{k})$ such that a loss of inner semi-continuity occurs a t_k. Hence, $(X^\star, \hat{t}\,)$ is an irregular accumulation point.

B1 if there exists $\varepsilon'' > 0$ such that $\mathcal{P}^\star(t)$ is continuous at any $t \in (\hat{t}-\varepsilon', \hat{t}+\varepsilon'')$ then, as for any $\delta > 0$ a continuous switch from unique to non-unique solutions must occur, we can construct a sequence of times $\{t_k\}_{k=1}^{\infty}$ at which a continuous bifurcation occurs converging to $\hat{t}$. Otherwise, we can proceed as in case **A2** and construct a sequence of times at which a loss of inner semi-continuity occurs converging to $\hat{t}$. Hence, $(X^\star, \hat{t}\,)$ is an irregular accumulation point.

B2 in this case, simply by definition, $(X^\star, \hat{t}\,)$ is a discontinuous isolated multiple point.

B3 if there exists $\varepsilon'' > 0$ such that $\mathcal{P}^\star(t)$ is continuous at any $t \in (\hat{t}+\varepsilon'')$, by definition $(X^\star, \hat{t}\,)$ is a discontinuous non-isolated multiple point (type 3.5). Otherwise, for every $k \in \mathbb{N}$ there exists a point $t_k \in (\hat{t}, \hat{t}+\frac{1}{k})$ such that a loss of inner semi-continuity occurs a t_k. Hence, $(X^\star, \hat{t}\,)$ is an irregular accumulation point.

B4 the same discussion as in **A3**, $(X^\star, \hat{t}\,)$ is hence an irregular accumulation point.

□

Theorem 3.4. *Let* $(\mathrm{P}_t, \mathrm{D}_t)$ *be a primal-dual pair of TV-SDPs parametrized over a time interval* $(0, \tau)$ *such that data* $\mathcal{A}_t, b_t, C_t$ *are polynomial functions of t and assumptions A1, A2, and A3, hold. Suppose that there exists a generic non-singular time.*

Then, along the parametrization interval $(0, \tau)$ *the number of points in times at which there is a non-differentiable point or a discontinuous isolated multiple point for* $\mathcal{P}^\star(t)$ *or* $\mathcal{D}^\star(t)$ *is finite. All the other points are regular points for both* $\mathcal{P}^\star(t)$ *and* $\mathcal{D}^\star(t)$*. Furthermore, the number of regular points where* $\mathcal{P}^\star(t)$ *or* $\mathcal{D}^\star(t)$ *is not continuously differentiable is finite.*

Proof. By Theorem 3.2, the hypothesis implies that the number of values of $t \in (0, \tau)$ at which there exists an optimal primal-dual singular point for (3.1) is finite. Let S denote the set of such values. First, let $\hat{t}_{ns} \in (0, \tau) \setminus S$. Then there exists an optimal primal-dual non-singular point $(X^\star_{ns}, Z^\star_{ns}, \hat{t}_{ns})$. By Theorem 3.1, both $(X^\star_{ns}, \hat{t}_{ns})$ and $(Z^\star_{ns}, \hat{t}_{ns})$

are regular points (cf. Def. 3.3 and Rem. 4) where both $\mathcal{P}^\star(t)$ and $\mathcal{D}^\star(t)$ are continuously differentiable. Now consider $\hat{t}_s \in S$. Then there exists an optimal primal-dual singular point $(X_s^\star, Z_s^\star, \hat{t}_s)$. If at $\hat{t}_s$ a loss of inner semi-continuity for $\mathcal{P}^\star$ occurs then $\mathcal{P}^\star(\hat{t}_s)$ is multi-valued, hence $(X_s^\star, \hat{t}_s)$ is a discontinuous isolated multiple point (cf. Def. 3.5). The same holds in the dual version for $\mathcal{D}^\star$ and $(Z_s^\star, \hat{t}_s)$. If instead at $\hat{t}_s$ continuity of $\mathcal{P}^\star$ is preserved, then $\mathcal{P}^\star(\hat{t}_s)$ is a singleton. According to whether $\mathcal{P}^\star$ is differentiable at $\hat{t}_s$ or not, $(X_s^\star, \hat{t}_s)$ is a regular point or a non-differentiable point (cf. Def. 3.4). At regular points in S that are differentiable, the derivative of $\mathcal{P}^\star(t)$ and $\mathcal{D}^\star(t)$ might yet fail to be continuous. Being in S, such points are in a finite number, hence proving the last sentence of the theorem. Since $\mathcal{P}^\star(\hat{t}_s)$ is a singleton, a loss of differentiability only happens when $\hat{t}_s$ is in S; that is, when either $\mathcal{D}^\star(\hat{t}_s)$ is multi-valued or strict complementarity between $X_s^\star$ and $Z_s^\star$ fails (this follows from Lemma 3.1). The same holds in the dual version for $\mathcal{D}^\star$ and $(Z_s^\star, \hat{t}_s)$. $\square$

Notice that while we only assume continuity of the data for Theorem 3.3, we need stronger regularity assumptions to guarantee the validity of Theorem 3.4 (as well as Theorems 3.1 and 3.2).

Other than interesting for a purely theoretical study, these results could be useful for algorithms design as follows. If one can guarantee that the conditions of Theorem 3.4 are satisfied, algorithms for time-varying optimization need not consider the behaviors corresponding to Definitions 3.6, 3.7, and 3.8. If, however, one would like to develop a solver for the case where only assumptions A1, A2, and A3 are satisfied, some rather pathological behaviors, such as non-isolated discontinuous multiple points or bifurcation points, need to be to considered. In this respect, we believe that our work has the merit of clarifying and making explicit the nature of the irregularities of the trajectories to TV-SDP. The precise algorithmic consequences will clearly be strongly dependent on the type and the properties of the algorithm in use. In the path-following algorithm that we propose in the next chapter, we will assume that the trajectory of the solutions is made by regular points, so that it describes a smooth and well-defined curve. In this chapter we clarified under which algebraic conditions such a regular behavior is ensured. Such a restrictive assumption is justified by Theorem 3.4, according to which the parametrization interval can be partitioned into open time subintervals made by regular point and a finite number of time singleton where regularity is lost.

To prove that any type of point that we defined can actually appear, in the following section we exhibit an example of each type.

Scene 3 Five exhaustive examples

Where we follow our heroes struggling to get out of the labyrinthine Singular Wood. Their vicissitudes are endless: after the unfortunate encounter with the Degenerate Witch, they have to face the spectre of count Cayley, a terrible ghost who has inhabited the forest for centuries. As if that were not enough, the constant bifurcations of the path make their attempt to find the way out of the wood futile, while misadventures keep accumulating on our heroes. Right when the two princes begin to lose hope, the exit from the dark forest suddenly appears bright in front of their eyes...

Example 1: regular, non-differentiable and discontinuous isolated multiple points

For $t \in (-3, 2)$, consider the primal TV-SDP

$$\begin{aligned} \min \quad & tx + ty + z \\ \text{s.t.} \quad & \begin{bmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{bmatrix} \succeq 0, \end{aligned} \tag{P_t^1}$$

whose feasible region is known as *Cayley spectrahedron*. We have:

$$\mathcal{P}^\star(t) = \begin{cases} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} & \text{for } t \in (-3, -2], \\[2ex] \begin{bmatrix} 1 & -t/2 & -t/2 \\ -t/2 & 1 & \frac{t^2}{2} - 1 \\ -t/2 & \frac{t^2}{2} - 1 & 1 \end{bmatrix} & \text{for } t \in (-2, 2) \setminus \{0\}, \\[2ex] \left\{ \begin{bmatrix} 1 & a & b \\ a & 1 & -1 \\ b & -1 & 1 \end{bmatrix} \;\middle|\; \begin{matrix} a + b = 0 \\ a, b \in [-1, 1] \end{matrix} \right\} & \text{at } t = 0. \end{cases}$$

In $(-3,-2)$, the trajectory is constant and all points are hence regular (Definition 3.3). In both intervals $(-2,0)$ and $(0,2)$, the solution to (P_t^1) is unique and the trajectory describes a parabolic differentiable curve and hence all its points are also regular.

Instead, $t=-2$ is a non-differentiable point (Definition 3.4). Indeed:

$$\frac{d}{dt}\mathcal{P}^\star(t)|_{t=-2^-} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & -0.5 & -0.5 \\ -0.5 & 0 & -2 \\ -0.5 & -2 & 0 \end{bmatrix} = \frac{d}{dt}\mathcal{P}^\star(t)|_{t=-2^+}.$$

Moreover, at $t=0$ there is a loss of uniqueness, as $\mathcal{P}^\star(0)$ is a one-dimensional face of Cayley spectrahedron. Thus, $t=0$ is a discontinuous isolated multiple point (Definition 3.5), as uniqueness is holding before for $t \in (-2,0)$ and after for $t \in (0,3)$.

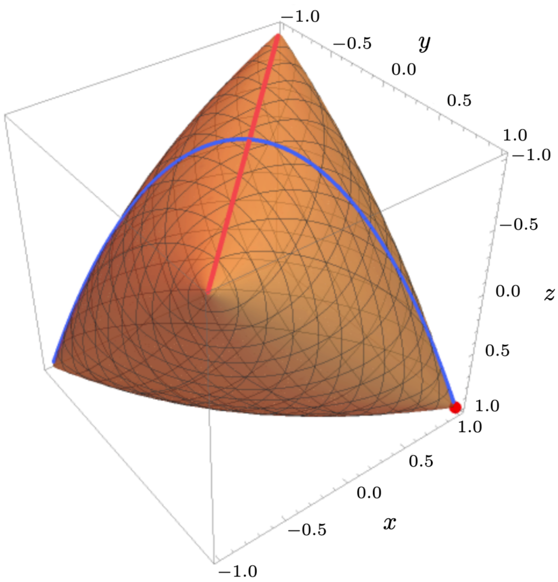


Figure 3.1: Trajectory of solutions of (P_t^1). Its feasible set is time-invariant and it is the Cayley spectrahedron (orange). Its optimal set-valued map coincides with the red dot at $(1,1,1)$ for $t \in (-3,-2]$, moves along the blue curve $(-t/2,-t/2,t^2/2-1)$ for $t \in (-2,2) \setminus \{0\}$, and covers the whole red top edge $\{(x,y,-1) | x+y=0\}$ at $t=0$.

Consider now the TV-SDP dual to (P_t^1):

$$\begin{aligned} \max \;\; & \alpha+\beta+\gamma \\ \text{s.t.} \;\; & \begin{bmatrix} -\alpha & t/2 & t/2 \\ t/2 & -\beta & 1/2 \\ t/2 & 1/2 & -\gamma \end{bmatrix} \succeq 0. \end{aligned} \tag{D_t^1}$$

The optimal set-valued map for (D_t^1) is

$$\mathcal{D}^\star(t) = \begin{cases} \begin{bmatrix} -t & t/2 & t/2 \\ t/2 & -(t+1)/2 & 1/2 \\ t/2 & 1/2 & -(t+1)/2 \end{bmatrix} & \text{for } t \in (-3,-2), \\ \\ \begin{bmatrix} t^2/2 & t/2 & t/2 \\ t/2 & 1/2 & 1/2 \\ t/2 & 1/2 & 1/2 \end{bmatrix} & \text{for } t \in [-2,2). \end{cases}$$

At $t=-2$, $\mathcal{D}^\star(t)$ has a non-differentiable point (Definition 3.4) too. Indeed:

$$\frac{d}{dt}\mathcal{D}^\star(t)|_{t=-2^-} = \begin{bmatrix} -1 & 0.5 & 0.5 \\ 0.5 & -0.5 & 0 \\ 0.5 & 0 & -0.5 \end{bmatrix} \neq \begin{bmatrix} -2 & 0.5 & 0.5 \\ 0.5 & 0 & 0 \\ 0.5 & 0 & 0 \end{bmatrix} = \frac{d}{dt}\mathcal{D}^\star(t)|_{t=-2^+}.$$

For $t \in (-3,2)\setminus\{-2\}$ the primal-dual pair of solutions is strictly complementary. Being both unique solutions for every $t \in (-3,2)\setminus\{0\}$, we conclude by Lemma 3.1 and Theorem 3.1 that for $t \in (-3,2)\setminus\{-2,0\}$ the primal-dual trajectory of solutions consists of regular points.

Notice that -2 and 0 are singular times for the parameterization interval $(-3,2)$. Indeed, at $t=-2$ there is a loss of strict complementarity (the rank of both primal and dual solution is 1), while at $t=0$ there is a loss of primal uniqueness, hence a dual degenerate solution.

Note that this example illustrates Theorem 3.4, as there exists a non-singular time $\hat{t} \in (-3,2)$ (Definition 3.2). Take for example $\hat{t}=1$: equation (3.1) has a finite set of 8 solutions, which can be described as the intersections in $\mathbb{R}^6$ of 3 sets, each of which is the union of 2 hyperplanes, with 3 hyperplanes. If we set

$$(X,Z) = \left(\begin{bmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{bmatrix}, \begin{bmatrix} -\alpha & 1/2 & 1/2 \\ 1/2 & -\beta & 1/2 \\ 1/2 & 1/2 & -\gamma \end{bmatrix} \right),$$

then equation (3.1) can be rewritten as:

$$\begin{cases} x = \alpha + \beta - \gamma, \\ y = \alpha - \beta + \gamma, \\ z = -\alpha + \beta + \gamma, \\ (1 + \alpha - \beta - \gamma)(1 + \beta + \gamma) = 0, \\ (1 - \alpha + \beta - \gamma)(1 + \alpha + \gamma) = 0, \\ (1 - \alpha - \beta + \gamma)(1 + \alpha + \beta) = 0. \end{cases} \tag{3.4}$$

The solutions of this system are:

$$\begin{array}{ll} (-\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}), & (1, 1, 1, 1, 1, 1), \\ (1, 1, -2, 1, -\tfrac{1}{2}, -\tfrac{1}{2}), & (-1, -1, 1, -1, 0, 0), \\ (1, -2, 1, -\tfrac{1}{2}, 1, -\tfrac{1}{2}), & (-1, 1, -1, 0, -1, 0), \\ (-2, 1, 1, -\tfrac{1}{2}, -\tfrac{1}{2}, 1), & (1, -1, -1, 0, 0, -1). \end{array}$$

It is possible to check that each of these eight points makes the Jacobian (3.3) invertible, hence guaranteeing that $\hat{t} = 1$ is a non-singular time, so that the hypothesis of Theorem 3.4 are satisfied. Notice that the first solution above corresponds to the optimal primal-dual solution at $\hat{t} = 1$.

Example 2: a discontinuous non-isolated multiple point

For $t \in (-2, 1)$, consider the TV-SDP

$$\begin{aligned} \min \quad & tx + ty + z \\ \text{s.t.} \quad & \begin{bmatrix} 1 & x & y & 0 \\ x & 1 & z & 0 \\ y & z & 1 & 0 \\ 0 & 0 & 0 & 1 + x + y + z \end{bmatrix} \succeq 0 \end{aligned} \tag{P_t^2}$$

for which

$$\mathcal{P}^\star(t) = \begin{cases} \left\{ \begin{bmatrix} 1 & -t/2 & -t/2 & 0 \\ -t/2 & 1 & \frac{t^2}{2}-1 & 0 \\ -t/2 & \frac{t^2}{2}-1 & 1 & 0 \\ 0 & 0 & 0 & \frac{t^2}{2}-t \end{bmatrix} \right\} & \text{for } t \in (-2,0), \\ \left\{ \begin{bmatrix} 1 & a & b & 0 \\ a & 1 & -1 & 0 \\ b & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \;\middle|\; \begin{matrix} a+b=0 \\ a,b \in [-1,1] \end{matrix} \right\} & \text{for } t \in [0,1). \end{cases}$$

The optimal set-valued map $\mathcal{P}^\star(t)$ is continuous for every $t \in (-2,1) \setminus \{0\}$, it is single-valued for every $t \in (-2,0)$, and it is multi-valued for every $t \in [0,1)$, as for every $t \in [0,1)$ the optimal face at t is one-dimensional. A loss of inner semincontinuity occurs at $t = 0$. Hence, $t = 0$ is a discontinuous non-isolated multiple point, according to Definition 3.6.

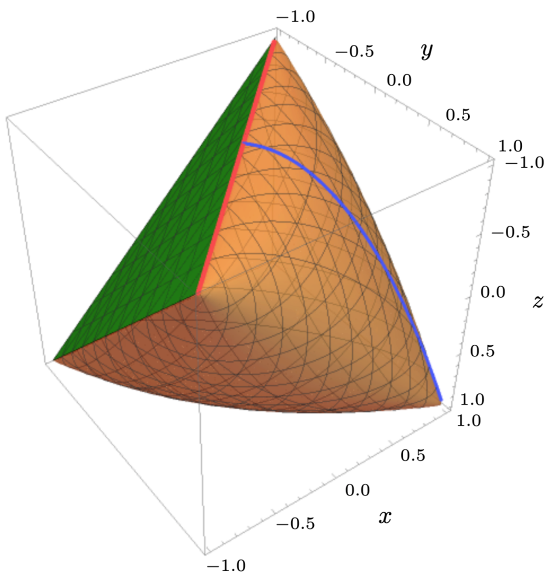


Figure 3.2: Trajectory of solutions of (P_t^2). Its feasible set is time-invariant and it is the Cayley spectrahedron (orange) intersected with half space $\{(x,y,z) | 1+x+y+z \geq 0\}$ (green). Its optimal set-valued map moves along the blue curve $(-t/2, -t/2, t^2/2-1)$ for $t \in (-1,0)$, and covers the whole red top edge $\{(x,y,z) | x+y=0, z=-1\}$ for $t \in [0,1)$.

Example 3: continuous bifurcation points

For $t \in (-1, 1)$, consider the primal TV-SDP

$$\begin{aligned} &\min X_{11} \\ &\text{s.t. } X_{44} - X_{33} = 0 \\ &\quad X_{22} = 1 \\ &\quad 2X_{12} + X_{33} + X_{44} = -t \\ &\quad X \succeq 0 \end{aligned} \tag{P_t^3}$$

for which

$$\mathcal{P}^\star(t) = \begin{cases} \left\{ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & a & b \\ 0 & a & -t/2 & c \\ 0 & b & c & -t/2 \end{bmatrix} \middle| \begin{array}{c} a^2+b^2+c^2 \le \frac{t^2}{4} - t \\ -\frac{t}{2}(a^2+b^2) + c^2 - 2abc \le \frac{t^2}{4} \end{array} \right\} & \text{for } t \in (-1, 0), \\ \begin{bmatrix} t^2/4 & -t/2 & 0 & 0 \\ -t/2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \text{for } t \in [0, 1). \end{cases}$$

The optimal set-valued map $\mathcal{P}^\star(t)$ is continuous for every $t \in (-1, 1)$, it is multi-valued for every $t \in (-1, 0)$, being there a three-dimensional face, and it is single-valued for every $t \in [0, 1)$ Hence $t = 0$ is a continuous bifurcation point for (P_t^3) according to Definition 3.7 (with reversed time, see Remark 6).

When there exists a continuous bifurcation point it is necessary that all the times of the parameterization interval are singular according to Definition 3.2. In other words, at any time $t \in (-1, 1)$ there exists a primal-dual point which is either degenerate or not strictly complementary.

Indeed, the dual TV-SDP to (P_t^3) is

$$\begin{aligned} \max \quad & y - tz \\ \text{s.t.} \quad & \begin{bmatrix} 1 & -z & 0 & 0 \\ -z & -y & 0 & 0 \\ 0 & 0 & -x-z & 0 \\ 0 & 0 & 0 & x-z \end{bmatrix} \succeq 0 \end{aligned} \tag{D_t^3}$$

which is equivalent to $\max\{y + tz \mid y + z^2 \leq 0,\ -z \leq x \leq z\}$ and for which

$$\mathcal{D}^\star(t) = \begin{cases} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \text{for } t \in (-1, 0], \\[2ex] \left\{ \begin{bmatrix} 1 & t/2 & 0 & 0 \\ t/2 & t^2/4 & 0 & 0 \\ 0 & 0 & -a & 0 \\ 0 & 0 & 0 & a+t \end{bmatrix} \middle|\ a \in [-t, 0] \right\} & \text{for } t \in (0, 1). \end{cases}$$

The dual optimal set-valued map $\mathcal{D}^\star(t)$ is continuous for every $t \in (-1, 1)$, single-valued for every $t \in (-1, 0]$, and it is multi-valued for every $t \in (0, 1)$, being there a one-dimensional face. Thus, $t = 0$ is a continuous bifurcation point for (D_t^3), according to Definition 3.7.

In particular, a pair of primal-dual solutions for (P_t^3, D_t^3) is not unique, hence degenerate, for every $t \in (-1, 1) \setminus \{0\}$. For $t = 0$, there is a unique pair of primal-dual solutions for which however strict complementarity does not hold. This implies that all $t \in (-1, 1)$ are singular times.

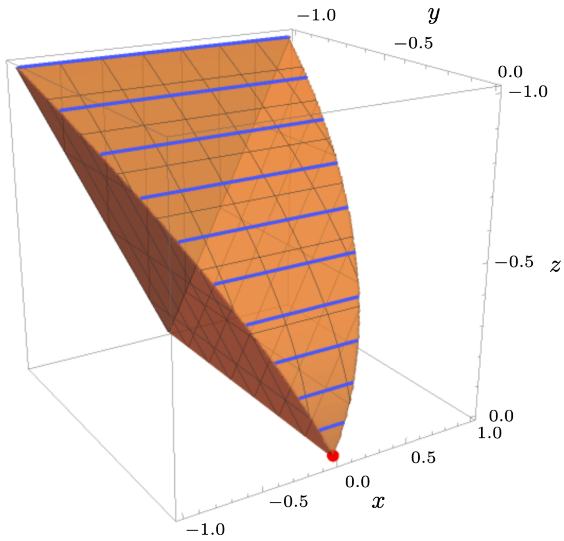


Figure 3.3: Trajectory of solutions of (D_t^3). Its feasible set is time-invariant and it is the set $\{(x,y,z) | y+z^2 \le 0,\ -z \le x \le z\}$ (orange). Its optimal set-valued map coincides with the red dot at $(0,0,0)$ for $t \in (-1,0]$. At $t=0$, $(0,0,0)$ is a continuous bifurcation point, as for every $t \in (0,1)$ the solution is multi-valued and equal to the set $\{(x,y,z) | x \in [-t/2, t/2], y = -t^2/4, z = -t/2\}$. In the picture, the blue segments illustrate the optimal multiple-valued solution for $t = \{0.1, 0.2, \ldots, 0.9, 1\}$

Example 4: a first irregular accumulation point

For $t \in (-1,1)$, consider the TV-SDP

$$\begin{aligned} \min\ & f(t)(x-y)+z \\ \text{s.t.}\ & \begin{bmatrix} 1 & x & y & 0 & 0 \\ x & 1 & z & 0 & 0 \\ y & z & 1 & 0 & 0 \\ 0 & 0 & 0 & g(t) & x-y \\ 0 & 0 & 0 & x-y & g(t) \end{bmatrix} \succeq 0 \end{aligned} \tag{P_t^4}$$

where

$$f(t) := \begin{cases} t \sin \frac{\pi}{t} & \text{if } t > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$g(t) := \begin{cases} 2t & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

For $t \leq 0$ the feasible region is the intersection between Cayley spectrahedron and the plane $x - y = 0$. For $t > 0$ the feasible region is the intersection between Cayley spectrahedron and the region $x - y \in [-2t, 2t]$.

Expressing the solutions of (P_t^4) in terms of the variables $x(t), y(t), z(t)$, we have:

$$(x(t), y(t), z(t)) = \begin{cases} (0,0,-1) & \text{for } t \in (-1, 0], \\ (t, -t, -1) & \text{for } t \in \left(\frac{1}{2k-1}, \frac{1}{2k}\right), \; k = 1, 2, \ldots \\ \{(\alpha, -\alpha, -1) \mid \alpha \in [-t, t]\} & \text{for } t = \frac{1}{k}, \qquad k = 1, 2, \ldots \\ (-t, t, -1) & \text{for } t \in \left(\frac{1}{2k}, \frac{1}{2k+1}\right), \; k = 1, 2, \ldots \end{cases}$$

For every $t \in (-1, 0]$, $\mathcal{P}^\star(t)$ is continuous and single-valued. The parameter sequence $\{t_k\}_{k=1}^\infty \subseteq (0, 1]$ defined by $t_k := \frac{1}{k}$ is such that $\lim_{k\to\infty} t_k = 0$ and at each t_k a loss of inner semi-continuity occurs. Hence, $t = 0$ is an irregular accumulation point, according to Definition 3.8

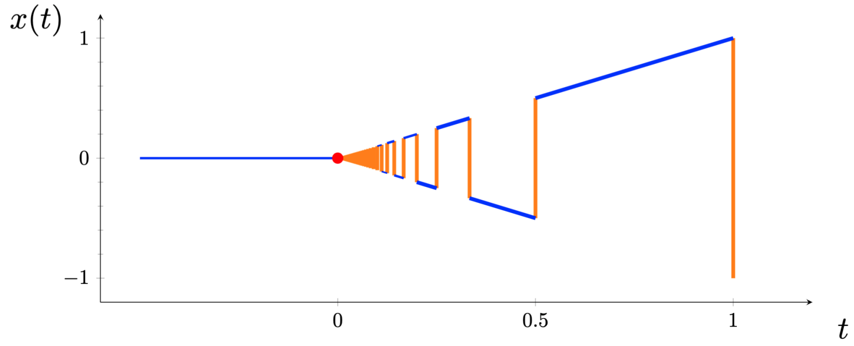


Figure 3.4: Graph of the x coordinate of the optimal set of (P_t^4) as a function of time t. The blue segments correspond to regular points, the red dot corresponds to an irregular accumulation point, and the orange vertical segments correspond to discontinuous isolated multiple-points, where the solution is multiple valued.

In the following, we also provide an example of an accumulation point for a sequence of continuous bifurcation points.

Example 5: a second irregular accumulation point

For $t \in (-1, 1)$, consider the TV-SDP

$$\begin{aligned} &\min \; z \\ &\text{s.t.} \begin{bmatrix} 1 & x & y & 0 & 0 \\ x & 1 & z & 0 & 0 \\ y & z & 1 & 0 & 0 \\ 0 & 0 & 0 & 2h(t) & x-y \\ 0 & 0 & 0 & x-y & 2h(t) \end{bmatrix} \succeq 0, \end{aligned} \tag{P$_t^5$}$$

where

$$h(t) := \begin{cases} t \sin^2 \frac{\pi}{t} & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

For $t \leq 0$ and for $t = 1/k,\ k = 1, 2, \ldots$ the feasible region is the intersection between Cayley spectrahedron and the plane $x - y = 0$, while for $t \in (1/k, 1/(k+1)),\ k = 1, 2, \ldots$ it is the intersection between Cayley spectrahedron and the region $x - y \in [-2h(t), 2h(t)]$.

Writing the solutions of (P$_t^5$) in terms of the variables $x(t), y(t), z(t)$, we have:

$$(x(t), y(t), z(t)) = \begin{cases} (0, 0, -1) & \text{for } t \in (-1, 0], \\ \{(\alpha, -\alpha, -1) \mid \alpha \in [-h(t), h(t)]\} & \text{for } t \in \left(\frac{1}{k}, \frac{1}{k+1}\right),\ k = 1, 2, \ldots \\ (0, 0, -1) & \text{for } t = \frac{1}{k}, \qquad k = 1, 2, \ldots \end{cases}$$

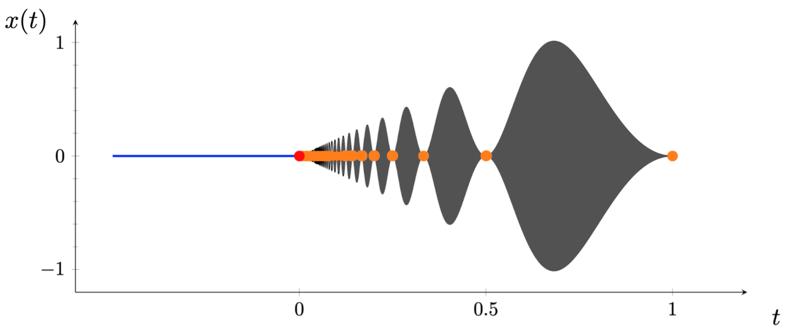
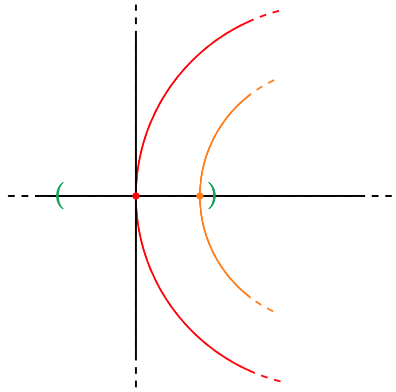
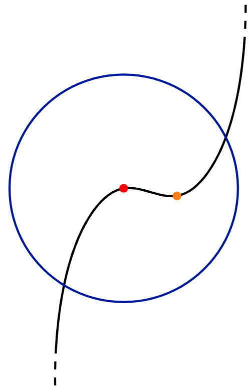


Figure 3.5: Graph of the x coordinate of the optimal set of (P$_t^5$) as a function of time t. The blue segment consists of regular points, the red dot corresponds to an irregular accumulation point, and the orange dots correspond to continuous bifurcation points. The gray region corresponds to times intervals where the optimal solution is multi-valued.

For every $t \in (-1, 1)$, $\mathcal{P}^\star(t)$ is continuous. The parameter sequence $\{t_k\}_{k=1}^{\infty} \subseteq (0, 1]$ defined by $t_k := \frac{1}{k}$ is such that $\lim_{k\to\infty} t_k = 0$ and each t_k is a continuous bifurcation point. Hence, $t = 0$ is an irregular accumulation point, according to Definition 3.8.

Chapter 4

Act III A path-following algorithm

"Quando me n'vo soletta per la via..."

Musetta in La Bohéme (Act III)

Let us start by recalling once again that we are considering TV-SDPs of the form

$$\begin{aligned} \min_{X \in \mathbb{S}^n} \quad & \langle C_t, X \rangle \\ \text{s.t.} \quad & \mathcal{A}_t(X) = b_t, \\ & X \succeq 0. \end{aligned} \tag{P$_t$}$$

Here $t \in [0, \tau]$ is a time parameter varying on a bounded and closed real interval. As usual, $\mathcal{A}_t : \mathbb{S}^n \to \mathbb{R}^m$ is a linear operator defined by $\mathcal{A}_t(X) = (\langle A_{1,t}, X \rangle, \dots, \langle A_{m,t}, X \rangle)$ for some $A_{1,t}, \dots, A_{m,t} \in \mathbb{S}^n$, $b_t \in \mathbb{R}^m$, and $C_t \in \mathbb{S}^n$. In this time-varying setting, one looks for a solution curve $t \mapsto X_t$ in $\mathbb{S}^n$ such that $X = X_t$ is an optimal solution for (P$_t$) at each time point $t \in [0, \tau]$.

A naive and immediate approach to solve the time-varying problem (P$_t$) is to consider, at a sequence of times $\{t_k\}_{k \in [\kappa]} \subseteq [0, \tau]$, the instances of the problem (P$_{t_k}$) for $k \in [\kappa]$ and solve them one after another. The best solvers for SDPs are interior point methods [10, 85, 86, 87, 88], which can solve them in a time that is polynomial in the input size. However, these solvers do not scale particularly well, and thus this brute-force approach may fail in applications where the volume and velocity of the data are large. Furthermore, such a straightforward method would not make use of the local information collected by solving the previous instances of the problem.

Unfortunately, IPMs are intrinsically unsuitable for warmstarting, lacking effective re-optimization strategies. In order to address this issue, various techniques where recently considered in the context of LP and SOCP [89, 90, 91, 92, 93] with discrete success, but is not clear whether these approaches can be efficiently extended to SDP.

Instead, in this work, we would like to utilize the idea of so-called path-following predictor-corrector algorithms as developed in [63, 64]. In classical predictor-corrector methods, a predictor step for approximating the directional derivative of the solution with respect to a small change in the time parameter is applied, together with a correction step that moves from the current approximate solution closer to the next solution at the new time point. The method that we propose incorporates these two steps in a single Newton step applied to the first-order optimality KKT conditions.

A limiting factor in solving both stationary and time-dependent SDPs is computational complexity when n is large. A common solution to this obstacle is the Burer–Monteiro approach, as presented in the seminal work [65, 66]. In this approach, a low-rank factorization $X = YY^T$ of the solution is assumed with $Y \in \mathbb{R}^{n\times r}$ and r potentially much smaller than n. In the optimization literature, the Burer–Monteiro method has been very well studied as a non-convex optimization problem, e.g. in terms of algorithms [68], quality of the optimal value [70, 71], and (global) recovery guarantees [72, 73, 74].

In a time-varying setting, the Burer–Monteiro factorization leads to

$$\begin{aligned} \min_{Y\in\mathbb{R}^{n\times r}} \quad & \langle C_t, YY^T\rangle \\ \text{s.t.} \quad & \mathcal{A}_t(YY^T) = b_t, \end{aligned} \tag{Q_t}$$

which for every fixed t is a quadratically constrained quadratic problem. A solution then is a curve $t \mapsto Y_t$ in $\mathbb{R}^{n\times r}$, which, depending on r, is a space of much smaller dimension than $\mathbb{S}^n$. However, this comes at the price that the problem (Q_t) is now non-convex. Moreover, theoretically it may happen that local optimization methods converge to a critical point that is not globally optimal [67], although in practice the method usually shows very good performance [65, 68, 69].

As we explain in the next section, to apply such methods, we need to address the issue that the solutions of (Q_t) are never isolated, due to the non-uniqueness of the Burer–Monteiro factorization caused by orthogonal invariance. We apply a well-known technique to handle this problem by restricting the solutions to a so-called *horizontal space* at every time step. From a geometric perspective, such an approach exploits the fact that equivalent factorizations can be identified as the same element in the corresponding quotient manifold with respect to the orthogonal group action [94].

Naturally, the rigorous formulation of path-following algorithms requires regularity assumptions on the solution curve. In our context, this will require both assumptions on the original TV-SDP problem (P_t) as well as on its reformulation (Q_t). In particular for the latter, the correct choice of the dimension r is crucial. In what follows, we present and discuss these assumptions in detail.

We recall that the dual problem of (P_t) is

$$\begin{aligned} \max_{y \in \mathbb{R}^m} \quad & \langle b_t, y \rangle \\ \text{s.t.} \quad & Z(y) := C_t - \mathcal{A}_t^*(y) \succeq 0 \end{aligned} \tag{D_t}$$

where $\mathcal{A}_t^* \colon y \mapsto \sum_{i=1}^m y_i A_{i,t}$ is the linear operator adjoint to $\mathcal{A}_t$. For convenience, we often drop the explicit dependence on y and refer to a solution of (D_t) simply as Z.

In this chapter, for the initial problem (P_t) we make the following assumptions.

(A1) Data $\mathcal{A}_t, b_t, C_t$ are continuously differentiable functions of t.

(A2) The linear operator $\mathcal{A}_t$ satisfies LICQ for any $t \in [0, \tau]$.

(A3) Problem (P_t) and its dual (D_t) are strictly feasible for every $t \in [0, \tau]$.

(A4) Problem (P_t) has a primal non-degenerate solution X_t and problem (D_t) has a dual non-degenerate solution Z_t at any $t \in [0, \tau]$.

(A5) The solution pair (X_t, Z_t) is strictly complementary for any $t \in [0, \tau]$.

Assumptions A1–5 rule out all the pathological behavior of the trajectory of the solutions to (P_t) exposed in the previous chapter, as well as in [61], so that all the points in the trajectory are regular. In particular, assumption A4 implies that the solution pair (X_t, Z_t) is unique. We recall that by Theorem 3.1 assumptions A1–5 have the following consequences:

(C1) Problem (P_t) has a unique and smooth solution curve X_t for every $t \in [0, \tau]$.

(C2) The curve $t \mapsto X_t$ is of constant rank $r^\star$.

For setting up the factorized version (Q_t) of (P_t), it is necessary to choose the dimension r of the factor matrix Y in (Q_t), ideally equal to $r^\star$ of C2. In what follows, we assume that we know the constant rank $r^\star$. Given access to an initial solution X_0 at time $t = 0$, it is possible to compute $r^\star$, so this assumption is without further loss of generality.

It is worth noting that the rank cannot be arbitrary: for any SDP defined by m linearly independent constraints, there always exists a solution of rank r such that $\frac{r(r+1)}{2} \leq m$. Since we assume that X_t is the unique solution to (P_t) with constant rank $r^\star$ we conclude that

$$r^\star \leq \left\lfloor \frac{\sqrt{1+8m}-1}{2} \right\rfloor.$$

The Barvinok–Pataki bound has been recently slightly improved by [95].

This chapter is structured as follows. In the first section we present the underlying quotient geometry of positive semidefinite rank-r matrices from a linear algebra perspective, focusing in particular on the notion of horizontal space and the domain of injectivity of the map $Y \mapsto YY^T$. We then present in the subsequent section our path-following algorithm, which is based on iteratively solving the linearized KKT system for (Q_t) over time. A main result there is the rigorous error analysis for this algorithm. Finally, in the last section, we showcase numerical results that test our method on a time-varying variant of the well-known Goemans–Williamson SDP relaxation for the Max-Cut problem in combinatorial optimization and graph theory.

Scene 1 Quotient geometry of positive semidefinite rank-r matrices

Where the two protagonists meet on the edge of the Singular Forest a low-ranking merchant gone into ruin: apparently, bad orbits messed up his horoscope, and he is now forced by hunger to beg for a loaf of bread. When the two generous princes offer him their help, he reveal to be no less than the mighty wizard of Burer–Monteiro, a land that is well-known to the princes, thanks to their cartography studies. The wizard rewards them by revealing the hidden location of the formidable Horizontal Spade, a magic weapon. Our heroes then continue their adventure with their new friend...

We now investigate the factorized formulation (Q_t) in more detail. As already mentioned, in contrast to the original problem (P_t), this is a nonlinear problem (specifically, a quadratically constrained quadratic problem) which is non-convex. Moreover, the property of uniqueness of a solution, which is guaranteed by C1 for the original problem (P_t), is lost in (Q_t), because its representation via the map

$$\phi : \mathbb{R}^{n\times r} \to \mathbb{S}^n, \quad \phi(Y) = YY^T$$

is not unique. In fact, this map is invariant under the orthogonal group action

$$\mathcal{O}_r \times \mathbb{R}^{n\times r} \to \mathbb{R}^{n\times r}, \quad (Q, Y) \mapsto YQ,$$

on $\mathbb{R}^{n\times r}$, where

$$\mathcal{O}_r := \{Q \in \mathbb{R}^{r\times r} \mid QQ^T = I_r\},$$

is the orthogonal group. Hence both the objective function $Y \mapsto \langle C_t, YY^T\rangle$ and the constraints $\mathcal{A}_t(YY^T) = b_t$ in (Q_t) are invariant under the same action. As a consequence, the solutions of (Q_t) are never isolated [68]. This poses a technical obstacle to the use of path-following algorithms, as the path needs to be, at least locally, uniquely defined.

On the other hand, by assuming that the correct rank $r = r^\star$ of a unique solution X_t for (P_t) has been chosen for the factorization, any solution Y_t for (Q_t) must satisfy $Y_t Y_t^T = X_t$. From this it follows that any solution is of the form $Y_t Q$ with $Q \in \mathcal{O}_r$; see, e.g., [66, Lemma 2.1]. In other words, the action of the orthogonal group is indeed the only source of non-uniqueness. This corresponds to the well-known fact that the set of positive definite fixed rank-r symmetric matrices, which we denote by $\mathcal{M}_r^+$, is a smooth manifold that can be identified with the quotient manifold $\mathbb{R}_*^{n\times r}/\mathcal{O}_r$, where $\mathbb{R}_*^{n\times r}$ is the open set of $n \times r$ matrices with full column rank.

In the following, we describe how the non-uniqueness can be removed by introducing the so-called horizontal spaces, which is a standard concept in optimization on quotient manifolds of the form [96]. For positive semidefinite fixed-rank matrices, this has been worked out in detail in [94]. Additional material, including the complex Hermitian case, can be found in [97]. However, in order to arrive at practical formulas that are useful for our path-following algorithm later on, we will not further refer to the concept of a quotient manifold but directly focus on the injectivity of the map ϕ on suitable linear subspaces of $\mathbb{R}^{n\times r}$, which we describe in this section. Such a simplification takes into account that we are dealing with a quotient manifold $\mathbb{R}_*^{n\times r}/\mathcal{O}_r$ with $\mathbb{R}_*^{n\times r}$ being just an open subset of $\mathbb{R}^{n\times r}$. Then the horizontal space at a point Y should be a subspace of the tangent space of $\mathbb{R}_*^{n\times r}$ at Y, which, however, is just $\mathbb{R}^{n\times r}$.

Given $Y \in \mathbb{R}_*^{n\times r}$, we denote the corresponding orbit under the orthogonal group as

$$Y\mathcal{O}_r := \{YQ \mid Q \in \mathcal{O}_r\} \subseteq \mathbb{R}_*^{n\times r}.$$

The orbit $Y\mathcal{O}_r$ is an embedded submanifold of $\mathbb{R}_*^{n\times r}$ of dimension T_{r-1} with two connected components, according to $\det Q = \pm 1$. Its tangent space at Y, which we denote by $\mathcal{T}_Y$, is easily derived by noting that the tangent space to the orthogonal group $\mathcal{O}_r$ at the identity matrix equals the space of real skew-symmetric matrices $\mathbb{S}^r_{skew}$ (see, e.g., [96, Example 3.5.3]). Therefore,

$$\mathcal{T}_Y = \{YS \mid S \in \mathbb{S}^r_{skew}\}.$$

Since the map $\phi(Y) = YY^T$ is constant on $Y\mathcal{O}_r$, its derivative

$$Y \mapsto \phi'(Y)[H] = YH^T + HY^T$$

vanishes on $\mathcal{T}_Y$, that is $\mathcal{T}_Y \subseteq \ker \phi'(Y)$.

The *horizontal space* at Y, denoted by $\mathcal{H}_Y$, is the orthogonal complement of $\mathcal{T}_Y$ with respect to the Frobenius inner product. One verifies that

$$\mathcal{H}_Y := \mathcal{T}_Y^\perp = \{H \in \mathbb{R}^{n\times r} \mid Y^T H = H^T Y\},$$

since $0 = \langle H, YS\rangle = \langle Y^T H, S\rangle$ holds for all skew-symmetric S if and only if $Y^T H$ is symmetric. We point out that sometimes any subspace complementary to $\mathcal{T}_Y$ is called a horizontal space, but we will stick to the above choice, as it is the most common and has certain theoretical and practical advantages. In particular, since $Y \in \mathcal{H}_Y$, the affine space $Y + \mathcal{H}_Y$ equals $\mathcal{H}_Y$, so it is just a linear space.

The purpose of the horizontal space is to provide a unique way of representing a neighborhood of $X = YY^T$ in $\mathcal{M}_r^+$ through $\phi(Y+H) = (Y+H)(Y+H)^T$ with $H \in \mathcal{H}_Y$. Clearly,

$$\dim \mathcal{H}_Y = nr - \dim \mathcal{O}_r = nr - T_{r-1} = \dim \mathcal{M}_r^+.$$

Moreover, the following holds.

Proposition 4.1. *The restriction of $\phi'(Y)$ to $\mathcal{H}_Y$ is injective. In particular, it holds that*

$$\|YH^T + HY^T\|_F \geq \sqrt{2}\sigma_r(Y)\|H\|_F \quad \textit{for all } H \in \mathcal{H}_Y,$$

where $\sigma_r(Y) > 0$ is the smallest singular value of Y. This lower bound is sharp if $r < n$. For $r = n$ one has the sharp estimate

$$\|YH^T + HY^T\|_F \geq 2\sigma_r(Y)\|H\|_F \quad \textit{for all } H \in \mathcal{H}_Y.$$

As a consequence, in either case, $\ker \phi'(Y) = \mathcal{T}_Y$.

Proof. For $Z \in \mathbb{S}^n$ we have $\operatorname{trace}((YH^T + HY^T)Z) = 2\operatorname{trace}(ZYH^T)$ by standard properties of the trace. Taking $Z = YH^T + HY^T$ yields

$$\|YH^T + HY^T\|_F^2 = 2\,\mathrm{trace}(YH^TYH^T + HY^TYH^T) = 2\|Y^TH\|_F^2 + 2\|YH^T\|_F^2.$$

To derive the second equality we used $Y^TH = H^TY$ for $H \in \mathcal{H}_Y$. Clearly, $\|YH^T\|_F^2 \geq \sigma_r(Y)^2\|H\|_F^2$ and if $r = n$ we also have that $\|Y^TH\|_F^2 \geq \sigma_r(Y)^2\|H\|_F^2$. This proves the asserted lower bounds. To show that they are sharp, let (u_r, v_r) be a (normalized) singular vector tuple such that $Yv_r = \sigma_r(Y)u_r$. If $r < n$, then for any u such that $u^TY = 0$ one verifies that the matrix $H = uv_r^T$ is in $\mathcal{H}_Y$ and achieves equality. When $r = n$, $H = u_r v_r^T$ achieves it. □

Since ϕ maps $\mathbb{R}_*^{n\times r}$ to $\mathcal{M}_r^+$, which is of the same dimension as $\mathcal{H}_Y$, the above proposition implies that $\phi'(Y)$ is a bijection between $\mathcal{H}_Y$ and $\mathcal{T}_{\phi(Y)}\mathcal{M}_r^+$. This already shows that the restriction of ϕ to the linear space $Y + \mathcal{H}_Y = \mathcal{H}_Y$ is a local diffeomorphism between a neighborhood of Y in $\mathcal{H}_Y$ and a neighborhood of $\phi(Y)$ in $\mathcal{M}_r^+$. The subsequent more quantitative statement matches the Theorem 6.3 in [94] on the injectivity radius of the quotient manifold $\mathbb{R}_*^{n\times r}/\mathcal{O}_r$. For convenience we will provide a self-contained proof that is more algebraic and does not require the concept of quotient manifolds.

Proposition 4.2. *Let $\mathcal{B}_Y := \{H \in \mathcal{H}_Y \mid \|H\|_F < \sigma_r(Y)\}$. Then the restriction of ϕ to $Y + \mathcal{B}_Y$ is injective and maps diffeomorphically to a (relatively) open neighborhood of Y in $\mathcal{M}_r^+$.*

It is interesting to note that $\mathcal{B}_Y$ is the largest possible ball in $\mathcal{H}_Y$ on which the result can hold, since the rank-one matrices $\sigma_i u_i v_i^T$ comprised of singular pairs of Y all belong to $\mathcal{H}_Y$ and $Y - \sigma_r u_r v_r^T$ is rank-deficient. Another important observation is that $\sigma_r(Y)$ does not depend on the particular choice of Y within the orbit $Y\mathcal{O}_r$.

Proof. Consider $H_1, H_2 \in \mathcal{B}_Y$. Let $Y = U\Sigma V^T$ be a singular value decomposition of Y with $U \in \mathbb{R}^{n\times r}$ and $V \in \mathbb{R}^{r\times r}$ having orthonormal columns. We assume $r < n$. Then by $U_\perp \in \mathbb{R}^{n\times(n-r)}$ we denote a matrix with orthonormal columns and $U^TU_\perp = 0$. In the case $r = n$, the terms involving $U_\perp$ in the following calculation are simply not present. We write

$$H_1 = UA_1V^T + U_\perp B_1 V^T, \quad H_2 = UA_2V^T + U_\perp B_2 V^T.$$

Since $H_1, H_2 \in \mathcal{H}_Y$, we have

$$\Sigma A_1 = A_1^T\Sigma, \quad \Sigma A_2 = A_2^T\Sigma.$$

Then a direct calculation yields

$$(Y+H_1)(Y+H_1)^T - YY^T = U[\Sigma A_1^T + A_1\Sigma + A_1A_1^T]U^T \\ + U[\Sigma + A_1]B_1^T U_\perp^T + U_\perp B_1[\Sigma + A_1^T]U^T + U_\perp B_1 B_1^T U_\perp^T,$$

and analogously for $(Y+H_2)(Y+H_2)^T - YY^T$. Since the four terms in the above sum are mutually orthogonal in the Frobenius inner product, the equality $(Y+H_1)(Y+H_1)^T = (Y+H_2)(Y+H_2)^T$ particularly implies

$$\Sigma A_1^T + A_1\Sigma + A_1A_1^T = \Sigma A_2^T + A_2\Sigma + A_2A_2^T,$$

as well as

$$(\Sigma + A_1)B_1^T = (\Sigma + A_2)B_2^T. \tag{4.1}$$

The first of these equations can be written as

$$\Sigma(A_1 - A_2)^T + (A_1 - A_2)\Sigma = A_2(A_2 - A_1)^T - (A_1 - A_2)A_1^T.$$

By Proposition 4.1 (with $n = r$, $Y = \Sigma$ and $H = A_1 - A_2$),

$$\|\Sigma(A_1 - A_2)^T + (A_1 - A_2)\Sigma\|_F \geq 2\sigma_r(Y)\|A_1 - A_2\|_F,$$

whereas

$$\|A_2(A_2 - A_1)^T - (A_1 - A_2)A_1^T\|_F \leq (\|H_2\|_F + \|H_1\|_F)\|A_1 - A_2\|_F.$$

Since $\|H_2\|_F + \|H_1\|_F < 2\sigma_r(Y)$, this shows that we must have $A_1 = A_2$, which then by (4.1) also implies $B_1 = B_2$, since $\Sigma + A_1$ is invertible.

Hence, we have proven that ϕ is an injective map from $Y + \mathcal{B}_Y$ to $\mathcal{M}_r^+$. To validate that it is a diffeomorphism onto its image we show that it is locally a diffeomorphism, for which again it suffices to confirm that $\phi'(Y+H)$ is injective on $\mathcal{H}_Y$ for every $H \in \mathcal{B}_Y$ (since $\mathcal{H}_Y$ and $\mathcal{M}_r^+$ have the same dimension). It follows from Proposition 4.1 (with Y replaced by $Y+H$, which has full column rank) that the null space of $\phi'(Y+H)$ equals $\mathcal{T}_{Y+H}$. We claim that $\mathcal{T}_{Y+H} \cap \mathcal{H}_Y = \{0\}$, which proves the injectivity of $\phi'(Y+H)$ on $\mathcal{H}_Y$. Indeed, let K be an element in the intersection, i.e., $K = (Y+H)S$ for some skew-symmetric S and $Y^TK - K^TY = 0$. Inserting the first relation into the second, and using $Y^TH = H^TY$, yields the homogenuous Lyapunov equation

$$(Y^TY + Y^TH)S + S(Y^TY + Y^TH) = 0. \tag{4.2}$$

The symmetric matrix

$$Y^TY + Y^TH = \frac{1}{2}(Y+H)^T(Y+H) + \frac{1}{2}(Y^TY - H^TH)$$

in (4.2) is positive definite, since $\lambda_1(H^TH) \leq \|H^TH\|_F < \sigma_r(Y)^2 = \lambda_r(Y^TY)$. But in this case (4.2) implies $S = 0$, that is, $K = 0$. □

Finally, it is also possible to provide a lower bound on the radius of the largest ball around $X = YY^T$ such that its intersection with $\mathcal{M}_r^+$ is in the image $\phi(Y + \mathcal{B}_Y)$, so that an inverse map ϕ^{-1} is defined.

Proposition 4.3. *Any $\tilde{X} \in \mathcal{M}_r^+$ satisfying $\|\tilde{X} - X\|_F < \frac{2\lambda_r(X)}{\sqrt{r+4}+\sqrt{r}}$ is in the image $\phi(Y + \mathcal{B}_Y)$, that is, there exists a unique $H \in \mathcal{B}_Y$ such that $\tilde{X} = (Y+H)(Y+H)^T$.*

Observe that one could take

$$\|\tilde{X} - X\|_F \leq \frac{\lambda_r(X)}{\sqrt{r+4}} \tag{4.3}$$

as a slightly cleaner sufficient condition in the proposition.

Proof. Let $\tilde{X} = \tilde{Z}\tilde{Z}^T$ with $\tilde{Z} \in \mathbb{R}^{n\times r}$ and assume a polar decomposition of $Y^T\tilde{Z} = P\tilde{Q}^T$, where $P, \tilde{Q} \in \mathbb{R}^{r\times r}$, P is positive semidefinite and $\tilde{Q}$ is orthogonal. Let $Z = \tilde{Z}\tilde{Q}$. Then

$$H = Z - Y \tag{4.4}$$

satisfies $(Y+H)(Y+H)^T = \tilde{X}$, and since $Y^TH = P - Y^TY$ is symmetric, we have $H \in \mathcal{H}_Y$. We need to show $H \in \mathcal{B}_Y$, that is, $\|H\|_F < \sigma_r(Y)$. Proposition 4.2 then implies that H is unique in $\mathcal{B}_Y$. Let $YY^\dagger$ be the orthogonal projector onto the column span of Y and $Z_1 = YY^\dagger Z$. With that, we have the decomposition

$$\|H\|_F^2 = \|YY^\dagger H\|_F^2 + \|(I - YY^\dagger)H\|_F^2 = \|Z_1 - Y\|_F^2 + \|(I - YY^\dagger)Z\|_F^2. \tag{4.5}$$

We estimate both terms separately. Since $Y^TZ_1 = Y^TZ = P$ is symmetric and positive semidefinite, the first term satisfies

$$
\begin{aligned}
\|Z_1 - Y\|_F^2 &= \|Z_1\|_F^2 - 2\,\mathrm{trace}(Y^T Z_1) + \|Y\|_F^2 \\
&= \|(Z_1 Z_1^T)^{1/2}\|_F^2 - 2\sum_{i=1}^{r} \sigma_i(Y^T Z_1) + \|(YY^T)^{1/2}\|_F^2. \qquad (4.6)
\end{aligned}
$$

A simple consideration using a singular value decomposition of Y and Z_1 reveals that

$$
(YY^T)^{1/2}(Z_1 Z_1^T)^{1/2} = \tilde{U} Y^T Z_1 \tilde{V}^T
$$

for some $\tilde{U}$ and $\tilde{V}$ with orthonormal columns. Consequently, by von Neumann's trace inequality (see, e.g., [23, Theorem 7.4.1.1]), we have

$$
\mathrm{trace}((YY^T)^{1/2}(Z_1 Z_1^T)^{1/2}) \le \sum_{i=1}^{r} \sigma_i(Y^T Z_1).
$$

Inserting this in (4.6) yields

$$
\|Z_1 - Y\|_F^2 \le \|(Z_1 Z_1^T)^{1/2} - (YY^T)^{1/2}\|_F^2.
$$

We remark that we could have concluded this inequality from [97, Theorem 2.7] where it is also stated. It actually holds for any Z_1 for which $Y^T Z_1$ is symmetric and positive semidefinite using the same argument (in particular for Z_1 replaced with the initial Z). Let now $Y = U\Sigma V^T$ be a singular value decomposition of Y with $\sigma_r(Y)$ the smallest positive singular value. Then $Z_1 Z_1^T = US^2U^T$ for some positive semidefinite $S^2 \in \mathbb{R}^{r\times r}$ and it follows from well-known results, cf. [98], that[1]

$$
\|(Z_1 Z_1^T)^{1/2} - (YY^T)^{1/2}\|_F^2 = \|S - \Sigma\|_F^2 \le \frac{1}{\sigma_r(Y)^2}\|S^2 - \Sigma^2\|_F^2 = \frac{1}{\sigma_r(Y)^2}\|Z_1 Z_1^T - YY^T\|_F^2.
$$

Noting that $Z_1 Z_1^T = (YY^\dagger)\tilde{X}(YY^\dagger)$ and $YY^T = (YY^\dagger)X(YY^\dagger)$ we conclude the first part with

$$
\|Z_1 - Y\|_F^2 \le \frac{1}{\sigma_r(Y)^2}\|(YY^\dagger)(\tilde{X} - X)(YY^\dagger)\|_F^2 \le \frac{1}{\sigma_r(Y)^2}\|\tilde{X} - X\|_F^2. \qquad (4.7)
$$

[1] For completeness we provide the proof. The matrix $S - \Sigma$ is the unique solution to the matrix equation $\mathcal{L}(M) = SM + M\Sigma = S^2 - \Sigma^2$. Indeed, the linear operator $\mathcal{L}$ on $\mathbb{R}^{r\times r}$ is symmetric in the Frobenius inner product and has positive eigenvalues $\lambda_{i,j} = \lambda_i(S) + \Sigma_{jj} \ge \sigma_r(Y)$ (the eigenvectors are rank-one matrices $w_i e_j^T$ with w_i the eigenvectors of S). Hence $\|S^2 - \Sigma^2\|_F = \|\mathcal{L}(S - \Sigma)\|_F \ge \sigma_r(Y)\|S - \Sigma\|_F$.

The second term in (4.5) can be estimated as follows:

$$
\begin{aligned}
\|(I-YY^{\dagger})Z\|_F^2 &= \operatorname{trace}((I-YY^{\dagger})\tilde{X}(I-YY^{\dagger})) \\
&\leq \sqrt{r}\|(I-YY^{\dagger})\tilde{X}(I-YY^{\dagger})\|_F \\
&= \sqrt{r}\|(I-YY^{\dagger})(\tilde{X}-X)(I-YY^{\dagger})\|_F \leq \sqrt{r}\|\tilde{X}-X\|_F, \qquad (4.8)
\end{aligned}
$$

where we used the Cauchy-Schwarz inequality and the fact that $(I-YY^{\dagger})\tilde{X}(I-YY^{\dagger})$ has rank at most r.

As a result, combining (4.5) with (4.7) and (4.8), we obtain

$$
\|H\|_F^2 \leq \frac{1}{\sigma_r(Y)^2}\|\tilde{X}-X\|_F^2 + \sqrt{r}\|\tilde{X}-X\|_F. \qquad (4.9)
$$

The right side is strictly smaller than $\sigma_r(Y)^2$ when

$$
\|\tilde{X}-X\|_F < -\frac{\sigma_r(Y)^2\sqrt{r}}{2} + \sqrt{\frac{\sigma_r(Y)^4 r}{4} + \sigma_r(Y)^4} = \frac{\sigma_r(Y)^2}{2}(\sqrt{r+4}-\sqrt{r}) = \frac{2\lambda_r(X)}{\sqrt{r+4}+\sqrt{r}},
$$

which proves the assertion. □

Remark 9. From definition (4.4) of H, since $\tilde{Q}$ is given by the polar decomposition $Y^T\tilde{Z} = P\tilde{Q}^T$, it follows that

$$
\|H\|_F = \|Y-\tilde{Z}\tilde{Q}\|_F = \min_{Q\in\mathcal{O}_r}\|Y-\tilde{Z}Q\|_F,
$$

see, e.g., [23, section 7.4.5]. In general, given any $Y,\tilde{Z} \in \mathbb{R}^{n\times r}$, both of rank r, the minimizer $Z = \tilde{Z}\tilde{Q}$ in this problem is necessarily obtained by choosing $\tilde{Q}$ from the polar decomposition of $Y^T\tilde{Z}$ so that Y^TZ is necessarily symmetric, that is, Z and hence $Z-Y$ are in the horizontal space $\mathcal{H}_Y$. In fact, the quantity $\min_{Q\in\mathcal{O}_r}\|Y-\tilde{Z}Q\|_F$ defines a Riemannian distance between the orbits $Y\mathcal{O}_r$ and $\tilde{Z}\mathcal{O}_r$ in the corresponding quotient manifold; see [94, Proposition 5.1].

We now return to the factorized problem formulation (Q_t). Let Y_t be an optimal solution of (Q_t) at some fixed time point t (so that $Y_tY_t^T = X_t$ and $\operatorname{rank} Y_t = r$). Based on the above propositions we are able to state a result on the allowed time interval $[t, t+\Delta t]$ for which the factorized problem (Q_t) is guaranteed to admit unique solutions on the horizontal space $\mathcal{H}_{Y_t}$ corresponding to the original problem (P_t). For this, exploiting the smoothness of the curve $t \mapsto X_t$, we first define

$$L := \max_{t\in[0,\tau]} \|\dot{X}_t\|_F, \tag{4.10}$$

a uniform bounds on the time derivative, as well as

$$\lambda_r(X_t) \geq \lambda_* > 0 \tag{4.11}$$

on the smallest eigenvalue of X_t, are available for $t \in [0,\tau]$. Notice that the existence of such bounds is without any further loss of generality: the existence of L follows from C1, which guarantees that X_t is a smooth curve, while the existence of λ_* is guaranteed by C2, since X_t has a constant rank.

Theorem 4.1. *Let Y_t be a solution of* (Q$_t$) *as above. Then for $\Delta t < \frac{2\lambda_*}{L(\sqrt{r+4}+\sqrt{r})}$ there is a unique and smooth solution curve $s \mapsto Y_s$ for the problem* (Q$_t$) *restricted to $\mathcal{H}_{Y_t}$ in the time interval $s \in [t, t+\Delta t]$.*

Proof. It suffices to show that for s in the asserted time interval the solutions X_s of (SDP$_s$) lie in the image $\phi(Y_t + \mathcal{B}_{Y_t})$. By Proposition 4.3, this is the case if $\|X_s - X_t\|_F < \frac{2\lambda_r(X_t)}{\sqrt{r+4}+\sqrt{r}}$. Since

$$\|X_s - X_t\|_F \leq \int_t^s \|\dot{X}_\tau\|_F \, d\tau \leq L(s-t),$$

and $\lambda_* \leq \lambda_r$, the condition $s - t < \frac{2\lambda_*}{L(\sqrt{r+4}+\sqrt{r})}$ is sufficient. Then Proposition 4.2 provides the smooth solution curve $Y_s = \phi^{-1}(X_s)$ for problem (Q$_t$). □

Similarly, if one wants to guarantee condition (4.3) it is sufficient to take

$$\Delta t < \frac{\lambda_*}{L\sqrt{r+4}} \tag{4.12}$$

The results of this section motivate the definition of a version of (Q$_t$) restricted to $\mathcal{H}_{Y_t}$, which we provide in the next section.

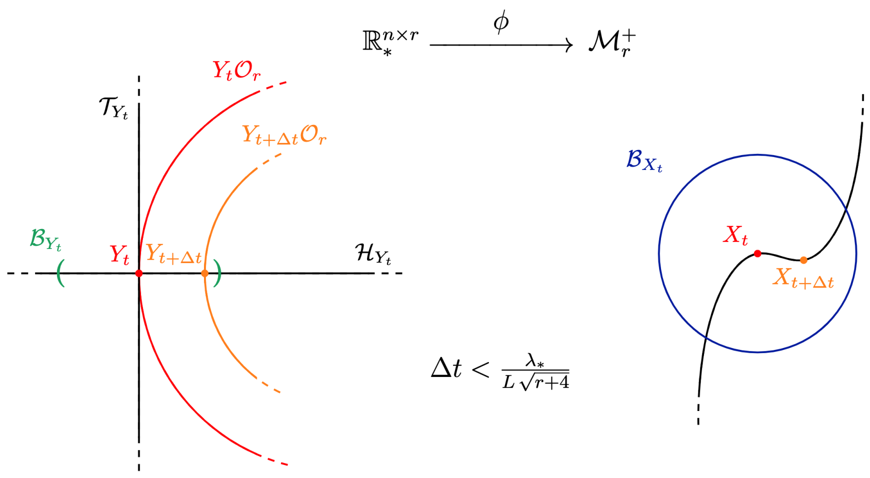


Figure 4.1: A visual summary of the results of the previous section: for Δt small enough, the solution $X_{t+\Delta t}$ is contained in the ball $\mathcal{B}_{X_t} := \{X \in \mathbb{S}^n \mid \|X - X_t\| \leq \lambda_r(X_t)/\sqrt{r+4}\}$, guaranteeing that there exists a unique matrix $Y_{t+\Delta t} \in \mathbb{R}^{n\times r}_*$ such that $\phi(Y_{t+\Delta t}) = X_{t+\Delta t}$ and such that $Y_{t+\Delta t} \in Y_t + \mathcal{B}_{Y_t} = \{H \in Y_t + \mathcal{H}_{Y_t} \mid \|H\|_F < \sigma_r(Y_t)\}$.

Scene 2 Path following the trajectory of solutions

Where our heroes begin the research of the Horizontal Spade. Unfortunately, in order to reach it they have to venture again in the Singular Wood. With the providential help of the wizard of Burer–Monteiro, they quickly find it in a clearing in the middle of the forest. Having taken possession of the enchanted weapon, they use its magic powers to make a clear and smooth way out the wood, which they follow step by step. This time the path errors are decisively bounded...

In this section, we present a path-following procedure for computing a sequence of approximate solutions $\{\hat{Y}_0, \ldots, \hat{Y}_k, \ldots, \hat{Y}_\kappa\}$ at different time points that tracks a trajectory of solutions $t \mapsto Y_t$ to the Burer–Monteiro reformulation (Q_t). From this sequence we are then able to reconstruct a corresponding sequence of approximate solutions $\hat{X}_k = \hat{Y}_k\hat{Y}_k^T$ tracking the trajectory of solutions $t \mapsto X_t$ for the full space TV-SDP problem (P_t). The path-following method is based on iteratively solving the linearized KKT system. Given an iterate Y_t on the path, we explained in the previous section how to eliminate the problem of non-uniqueness of the path in a small time interval $[t, t + \Delta t]$ by considering problem (Q_t) restricted to the horizontal space $\mathcal{H}_{Y_t}$. We now need to ensure that this also guarantees that the linearized KKT system admits a unique solution. We show in Theorem 4.2 that this is indeed guaranteed under standard regularity assumptions on the original problem (P_t). This is a remarkable fact of somewhat independent interest.

Linearized KKT conditions and second-order sufficiency

Given an optimal solution $X_t = Y_t Y_t^T$ at time t, we aim to find a solution $X_{t+\Delta t} = Y_{t+\Delta t} Y_{t+\Delta t}^T$ at time $t + \Delta t$. By the results of the previous section, the next solution can be expressed in a unique way as

$$Y_{t+\Delta t} = Y_t + \Delta Y,$$

where ΔY is in the horizontal space $\mathcal{H}_{Y_t}$, provided that Δt is small enough.

We define the following maps:

$$\begin{aligned} f_{t+\Delta t}(Y) &:= \langle C_{t+\Delta t}, YY^T \rangle, \\ g_{t+\Delta t}(Y) &:= \mathcal{A}_{t+\Delta t}(YY^T) - b_{t+\Delta t}, \\ h_{Y_t}(Y) &:= Y_t^T Y - Y^T Y_t. \end{aligned} \tag{4.13}$$

By definition, $\Delta Y \in \mathcal{H}_{Y_t}$ if and only if $h_{Y_t}(\Delta Y) = 0$. For symmetry reasons we use the equivalent condition $h_{Y_t}(Y_t + \Delta Y) = 0$ (which reflects the fact that $Y_t + \mathcal{H}_{Y_t}$ is actually a linear space).

To find the new iterate $Y_{t+\Delta t}$ we hence consider the problem

$$\begin{aligned} \min_{Y \in \mathbb{R}^{n \times r}} \quad & f_{t+\Delta t}(Y) \\ \text{s.t.} \quad & g_{t+\Delta t}(Y) = 0 \\ & h_{Y_t}(Y) = 0. \end{aligned} \tag{$\mathrm{Q}_{Y_t,t+\Delta t}$}$$

This is a quadratically constrained quadratic problem whose Lagrangian is

$$\mathcal{L}_{Y_t,t+\Delta t}(Y, \mu, \nu) := f_{t+\Delta t}(Y) - \langle \mu, g_{t+\Delta t}(Y) \rangle - \langle \nu, h_{Y_t}(Y) \rangle \tag{4.14}$$

with multipliers $\mu \in \mathbb{R}^m$ and $\nu \in \mathbb{S}^r_{skew}$. The KKT conditions of problem ($\mathrm{Q}_{Y_t,t+\Delta t}$) are

$$\begin{aligned} \nabla_Y \mathcal{L}_{Y_t,t+\Delta t}(Y, \mu, \nu) &= 0 \\ g_{t+\Delta t}(Y) &= 0 \\ h_{Y_t}(Y) &= 0. \end{aligned} \tag{4.15}$$

Hence, equation (4.15) reads explicitly as

$$\mathcal{F}_{Y_t,t+\Delta t}(Y,\mu,\nu) := \begin{bmatrix} 2C_{t+\Delta t}Y - 2\mathcal{A}^*_{t+\Delta t}(\mu)Y - 2Y_t\nu \\ \mathcal{A}_{t+\Delta t}(YY^T) - b_{t+\Delta t} \\ Y_t^TY - Y^TY_t \end{bmatrix} = 0.$$

The linearization of (4.15) at (Y_t, μ_t, ν_t) leads to a linear system

$$\mathcal{J}_{Y_t,t+\Delta t}(Y_t,\mu_t,\nu_t) \begin{bmatrix} \Delta Y \\ \mu_t + \Delta\mu \\ \nu_t + \Delta\nu \end{bmatrix} = \begin{bmatrix} -\nabla_Y f_{t+\Delta t}(Y_t) \\ g_{t+\Delta t}(Y_t) \\ 0 \end{bmatrix}, \tag{L}$$

where $\mathcal{J}_{Y_t,t+\Delta t}(Y,\mu,\nu)$ denotes the derivative of $\mathcal{F}_{Y_t,t+\Delta t}$ at (Y,μ,ν). Note that it actually does not depend on ν, but we will keep this notation for consistency. As a linear operator on $\mathbb{R}^{n\times r}\times\mathbb{R}^m\times\mathbb{S}^r_{skew}$, $\mathcal{J}_{Y_t,t+\Delta t}(Y,\mu,\nu)$ can be written in block matrix notation as follows:

$$\mathcal{J}_{Y_t,t+\Delta t}(Y,\mu,\nu) := \begin{bmatrix} \nabla^2_Y\mathcal{L}_{Y_t,t+\Delta t}(\mu) & -g'_{t+\Delta t}(Y)^* & -h^*_{Y_t} \\ -g'_{t+\Delta t}(Y) & 0 & 0 \\ -h_{Y_t} & 0 & 0 \end{bmatrix},$$

where from (4.13) and (4.14) one derives

$$\begin{aligned} \nabla^2_Y\mathcal{L}_{Y_t,t+\Delta t} &: H \mapsto 2(C_{t+\Delta t} - \mathcal{A}^*_{t+\Delta t}(\mu))H, \\ g'_{t+\Delta t}(Y) &: H \mapsto \mathcal{A}_{t+\Delta t}(YH^T + HY^T), \\ h_{Y_t} &: H \mapsto Y_t^TH - H^TY_t, \\ g'_{t+\Delta t}(Y)^* &: \mu \mapsto 2\mathcal{A}^*_{t+\Delta t}(\mu)Y, \\ h^*_{Y_t} &: \nu \mapsto 2Y_t\nu. \end{aligned}$$

For later reference, observe that as a bilinear form $\nabla^2_Y\mathcal{L}_{Y_t,t+\Delta t}$ reads

$$\nabla^2_Y\mathcal{L}_{Y_t,t+\Delta t}(\mu)[H,H] = 2\,\mathrm{trace}(H^T(C_{t+\Delta t} - \mathcal{A}^*_{t+\Delta t}(\mu))H).$$

Solving system (L) to obtain updates $(Y_t + \Delta Y, \mu_t + \Delta\mu, \nu_t + \Delta\nu)$ is equivalent to applying one step of Newton's method to the KKT system (4.15) (Lagrange–Newton method).

Our aim in this subsection is to show that for Δt small enough the system (L) is uniquely solvable when (Y_t, μ_t) is a KKT-pair for the overparametrized problem (Q_t). Since the system is continuous in Δt, we can do that by showing that it admits a unique solution for $\Delta t = 0$. This corresponds to proving second-order sufficient conditions for the optimality of problem ($Q_{Y_t, t+\Delta t}$) for $\Delta t = 0$. Interestingly, it is possible to relate this to standard regularity hypotheses on the original semidefinite problem (P_t). For this we first need a uniqueness statement on the Lagrange multiplier μ_t.

Lemma 4.1. *Given an optimal solution $X_t = Y_t Y_t^T$ to (P_t), suppose that X_t is a unique (see consequence C1), primal non-degenerate (see Definition 2.9 and assumption A4) solution. Then there is a unique optimal Lagrangian multiplier μ_t for (Q_t) independent from the choice of Y_t in the orbit $Y_t\mathcal{O}_r$. Moreover, $Z(\mu_t) = C_t - \mathcal{A}_t^*(\mu_t)$ is the unique dual solution to (D_t).*

Proof. We start by recalling that the optimal set for (Q_t) coincide with $Y_t\mathcal{O}_r$. Since the KKT conditions for (Q_t) are just

$$\nabla_Y f_t(Y) - \nabla_Y \langle \mu, g_t(Y) \rangle = 2(C_t - \mathcal{A}_t^*(\mu))Y = 0$$

(and $g_t(Y) = 0$), the set of all the optimal dual multipliers for (Q_t) is given by the set

$$\{\mu \mid (C_t - \mathcal{A}_t^*(\mu))Y_t Q = 0, Q \in \mathcal{O}_r\} = \{\mu \mid (C_t - \mathcal{A}_t^*(\mu))Y_t = 0\}$$

To show that this set is a singleton, it then suffices to prove that the homogeneous equation $\mathcal{A}_t^*(\mu)Y_t = 0$ has only the zero solution. By (2.4), primal non-degeneracy for X_t can read as

$$\operatorname{im} \mathcal{A}_t^* \cap \mathcal{T}_{X_t}^\perp = \{0\},$$

where $\mathcal{T}_{X_t}^\perp = \{M \in \mathbb{S}^n \mid MX_t = 0\}$. Noticing that $\mathcal{A}_t^*(\mu)Y_t = 0$ implies $\mathcal{A}_t^*(\mu) \in \operatorname{im}(\mathcal{A}_t^*) \cap \mathcal{T}_{X_t}^\perp$, we get that $\mathcal{A}_t^*(\mu) = 0$ and thus $\mu = 0$ since $\mathcal{A}_t^*$ is injective by assumption A2. To prove the second statement, observe that by primal non-degeneracy the dual problem (D_t) has a unique solution $Z(y_t)$ corresponding, by assumption A2, to a unique dual multipliers vector y_t (see Theorem 7 in [38]). Furthermore, $Z(y_t)$ satisfies $Z(y_t)X_t = \left(C_t - \mathcal{A}_t^*(y_t)\right) Y_t Y_t^T = 0$ by (KKT). Since Y_t has full column rank if r is chosen equal to $r^\star = \operatorname{rank} X_t$, this implies that $\left(C_t - \mathcal{A}_t^*(y_t)\right) Y_t = 0$. From the first statement it then follows that $y_t = \mu_t$. □

We can now state and prove the main result of this subsection.

Theorem 4.2. *Let $(X_t = Y_t Y_t^T, Z_t)$ be an optimal primal-dual pair of solutions to $(\mathrm{P}_t, \mathrm{D}_t)$ which is strictly complementary (see Definition 2.8) and such that X_t is primal non-degenerate (see Definition 2.9). Let μ_t be the unique corresponding Lagrange multiplier for (Q_t) according to Lemma 4.1. Then the triple $(Y_t, \mu_t, \nu_t = 0)$ is a KKT triple for $(\mathrm{Q}_{Y_t, t+\Delta t})$ at $\Delta t = 0$ (that is, $\mathcal{F}_{Y_t,t}(Y_t, \mu_t, 0) = 0$) and fulfills the SOSC:*

$$\nabla_Y^2 \mathcal{L}_{Y_t,t}(\mu_t)[H,H] = \mathrm{trace}(H^T(C_t - \mathcal{A}_t^*(\mu_t))H) > 0 \tag{4.16}$$

for all $H \in \mathbb{R}^{n\times r} \setminus \{0\}$ satisfying $\mathcal{A}_t(Y_t H^T + H Y_t^T) = 0$ and $Y_t^T H - H Y_t^T = 0$. In particular, $\mathcal{J}_{Y_t,t}(Y_t, \mu_t, 0)$ is invertible.

Proof. Since $(C_t - \mathcal{A}^*(\mu_t))Y_t = Z(\mu_t)Y_t = 0$ by the KKT conditions for (Q_t) and $h_{Y_t}(Y_t) = 0$, it is obvious that $\mathcal{F}_{Y_t,t}(Y_t, \mu_t, 0) = 0$. It is well-known that the linearized KKT system (L) admits a unique solution if (and only if) the SOSC (4.16) hold; see e.g., [99, Lemma 16.1]. Since $(X_t, Z(\mu_t))$ is an optimal solution for the original primal-dual pair of SDPs, and it hence satisify the second-order necessary conditions for optimality (that is, $Z(\mu_t) \succeq 0$), (4.16) holds with "$\geq$". Assume that

$$\mathrm{trace}(H^T(C_t - \mathcal{A}_t^*(\mu_t))H) = \mathrm{trace}(H^T Z(\mu_t)H) = 0$$

for some $H \in \mathbb{R}^{n\times r}$ satisfying $\mathcal{A}_t(Y_t H^T + H Y_t^T) = 0$ and $Y_t^T H - H^T Y_t = 0$. Since $Z_t = Z(\mu_t)$ is positive semidefinite, the columns of H must belong to the kernel of $Z(\mu_t)$. By strict complementarity (2.3) they hence belong to the column space of X_t, which is equal to the column space of Y_t. Therefore $H = Y_t P$ for some matrix $P \in \mathbb{R}^{r\times r}$. Consider now the matrix

$$\tilde{X} = X_t + s(Y_t H^T + H Y_t^T) = Y_t[I_r + s(P^T + P)]Y_t^T,$$

depending on a real parameter s. Clearly, $\mathcal{A}_t(\tilde{X}) = b_t$ and, for non-zero $|s|$ small enough, $\tilde{X}$ is positive semidefinite. Furthermore, for a suitable choice of the sign of s, we have $\langle C_t, \tilde{X}\rangle \leq \langle C_t, X_t\rangle$. Since X_t is the unique solution of (P_t), this implies $\tilde{X} = X_t$ and thus $Y_t H^T + H Y_t^T$ must be zero. Since $H \in \mathcal{H}_{Y_t}$ Proposition 4.1 yields $H = 0$, and this completes the proof. □

Corollary 4.1. *Let the assumptions of Theorem 4.2 be satisfied. Then for $\Delta t > 0$ small enough (and depending on Y_t) system (L), that is, operator $\mathcal{J}_{Y_t,t+\Delta t}(Y_t, \mu_t, 0)$ is invertible.*

Clearly, this is only a qualitative result. An upper bound for feasible Δt could be expressed in terms of the spectral norm of the inverse of $\mathcal{J}_{Y_t,t}(Y_t, \mu_t, 0)$ using perturbation arguments. This would require a lower bound on the absolute value of the eigenvalues of $\mathcal{J}_{Y_t,t}(Y_t,\mu_t,0)$. In this context, we should clarify that the eigenvalues, and hence also the condition number of $\mathcal{J}_{Y_t,t+\Delta t}(Y_t,\mu_t,0)$ (for sufficiently small Δt as above), do not depend on the particular choice of Y_t in the orbit $Y_t\mathcal{O}_r$. This is obviously also relevant from a practical perspective. To see this, note that as a bilinear form on $\mathbb{R}^{n\times r}\times\mathbb{R}^m\times\mathbb{S}^r_{skew}$ $\mathcal{J}_{Y_t,t+\Delta t}(Y,\mu,\nu)$ reads

$$\begin{aligned}
&\mathcal{J}_{Y_t,t+\Delta t}(Y,\mu,\nu)[(H,\Delta\mu,\Delta\nu),(H,\Delta\mu,\Delta\nu)] \qquad (4.17)\\
&= \operatorname{trace}(H^T(C_{t+\Delta t}-\mathcal{A}^*_{t+\Delta t}(\mu))H)-2\langle\Delta\mu,\mathcal{A}_{t+\Delta t}(YH^T+HY^T)\rangle-2\langle\Delta\nu,Y_t^TH-H^TY_t\rangle.
\end{aligned}$$

For any fixed $Q\in\mathcal{O}_r$ one therefore has

$$\begin{aligned}
&\mathcal{J}_{Y_t,t+\Delta t}(Y_t,\mu_t,0)[(H,\Delta\mu,\Delta\nu),(H,\Delta\mu,\Delta\nu)]\\
&=\mathcal{J}_{Y_tQ,t+\Delta t}(Y_tQ,\mu_t,0)[\mathcal{T}_Q(H,\Delta\mu,\Delta\nu),\mathcal{T}_Q(H,\Delta\mu,\Delta\nu)]
\end{aligned}$$

with the unitary linear operator $\mathcal{T}_Q(H,\Delta\mu,\Delta\nu)=(HQ,\Delta\mu,Q^T\Delta\nu Q)$ on $\mathbb{R}^{n\times r}\times\mathbb{R}^m\times\mathbb{S}^r_{skew}$. It follows that $\mathcal{J}_{Y_t,t+\Delta t}(Y_t,\mu_t,0)$ and $\mathcal{J}_{Y_tQ,t+\Delta t}(Y_tQ,\mu_t,0)$ have the same eigenvalues.

However, our proof of Theorem 4.2 is by contradiction and hence does not provide an obvious lower bound on the radius of invertibility of $\mathcal{J}_{Y_t,t}(Y_t,\mu_t,0)$. Here we do not intend to investigate this in more depth. In the error analysis conducted later we will essentially assume to have such a bound available (cf. Lemma 4.3).

A path-following algorithm

We now thoroughly describe the path-following algorithm that we propose for tracking the trajectory of solutions to (P_t). It includes an optional adaptive stepsize tuning step which is based on measuring the *residual* of the optimality conditions, defined as

$$\operatorname{res}_t(Y,\mu):=\left\|\begin{array}{c}2[C_t-\mathcal{A}_t^*(\mu)]Y\\ \mathcal{A}_t(YY^T)-b_t\end{array}\right\|_\infty. \qquad \text{(RES)}$$

The residual expresses the maximal component-wise violation of the optimality KKT conditions for the problem (Q_t) and is therefore a suitable error measure. Indeed (see, e.g., [100, Theorems 3.1 and 3.2]), if the second-order sufficiency condition for optimality holds at (Y_t,μ_t), then there are constants $\eta,C_1,C_2>0$ such that for all (Y,μ) with

$\|(Y,\mu)-(Y_t,\mu_t)\| \leq \eta$ one has

$$C_1\|(Y,\mu)-(Y_t,\mu_t)\| \leq \operatorname{res}_t(Y,\mu) \leq C_2\|(Y,\mu)-(Y_t,\mu_t)\|.$$

Here and in the following, we we use the norm $\|(Y,\mu)\|^2 = \|Y\|_F^2 + \|\mu\|^2$.

Algorithm 1 A primal-dual path-following algorithm for (P_t, D_t) with $t \in [0,\tau]$

Inputs: an initial approximate primal-dual solution $(\hat{X}_0, Z(\hat{\mu}_0))$ for $t=0$
initial stepsize Δt_0
boolean variable `STEPSIZE_TUNING`
stepsize tuning parameters $\gamma_1 \in (0,1)$, $\gamma_2 > 1$
residual tolerance $\epsilon > 0$

Output: list of approximate primal-dual solutions $\{(\hat{X}_k, Z(\hat{\mu}_k))\}_{k\in[\kappa]}$ corresponding to a partition $\{t_k\}_{k\in[\kappa]}$ of size κ of the interval $[0,\tau]$.

1: $k \leftarrow 0$
2: $t_0 \leftarrow 0$
3: $\Delta t \leftarrow \Delta t_0$
4: $S = \{(\hat{X}_0, Z(\hat{\mu}_0))\}$
5: $r = \operatorname{rank}(\hat{X}_0)$
6: find $\hat{Y}_0 \in \mathbb{R}^{n\times r}$ such that $\hat{Y}_0\hat{Y}_0^T = \hat{X}_0$
7: **while** $t_k < \tau$ **do**
8: solve linear system (L) with data $\Delta t, t_k, \hat{Y}_k, \hat{\mu}_k$ and obtain $\Delta Y, \Delta\mu$
9: **if** `STEPSIZE_TUNING` **and** $\operatorname{res}_{\hat{Y}_k, t_k+\Delta t}(\hat{Y}_k+\Delta Y, \hat{\mu}_k+\Delta\mu) > \epsilon$ **then**
10: $\Delta t \leftarrow \gamma_1 \Delta t$
11: go back to step 7
12: $(t_{k+1}, \hat{Y}_{k+1}, \hat{\mu}_{k+1}) \leftarrow (t_k+\Delta t, \hat{Y}_k+\Delta Y, \hat{\mu}_k+\Delta\mu)$
13: append $(\hat{X}_{k+1} = \hat{Y}_{k+1}\hat{Y}_{k+1}^T, Z(\hat{\mu}_{k+1}))$ to S
14: **if** `STEPSIZE_TUNING` **then**
15: $\Delta t \leftarrow \min(\tau - t_{k+1}, \gamma_2\Delta t, \Delta t_0)$
16: **else**
17: $\Delta t \leftarrow \min(\tau - t_{k+1}, \Delta t)$
18: $k \leftarrow k+1$
19: **return** S

The overall procedure is displayed as Algorithm 1 above. Given a TV-SDP of the form (P_t), parameterized over a time interval $[0, \tau]$, the inputs are an approximate initial primal-dual solution pair $(\hat{X}_0, Z(\hat{\mu}_0))$ to ($\mathrm{P}_0, \mathrm{D}_0$) and an initial stepsize Δt_0. At each iteration the current iterate is used to construct the linear system (L), which is then solved, returning the updates ΔY and $\Delta \mu$. The presented version of the algorithm also includes a procedure for tuning the stepsize that can be activated through the Boolean variable `STEPSIZE_TUNING` and is supposed to ensure that the residual threshold is satisfied at every time step. Specifically, if for a time step the threshold is violated, the stepsize is reduced by a factor $\gamma_1 \in (0,1)$ and a more accurate solution is obtained by solving the linearized KKT system (L) for the reduced time step. On the other hand, for avoiding unnecessary small steps, the stepsize is increased after every successful step by a factor $\gamma_2 > 1$ (but is never made larger than Δt_0). If the stepsize tuning is deactivated, the algorithm just runs with the constant stepsize Δt_0 instead.

Note that the algorithm tracks both the primal solution X_t and the dual solution $Z_t = C_t - \mathcal{A}_t^*(\mu_t)$.

Error analysis

We investigate the algorithm without stepsize tuning. The main goal of the following error analysis to show that the computed $(\hat{X}_k, \hat{\mu}_k)$, where $\hat{X}_k = \hat{Y}_k \hat{Y}_k^T$, remain close to the exact solutions (X_{t_k}, μ_{t_k}), if properly initialized. The logic of the proof is similar to standard path following methods based on Newton's method, e.g. [101]. The specific form of our problem requires some additional considerations that allow for more precise quantitative bounds depending on the problem constants.

Throughout this section, $(X_t = Y_t Y_t^T, Z_t)$ is an optimal primal-dual pair of solutions to ($\mathrm{P}_t, \mathrm{D}_t$) satisfying assumptions A1–5, so that it is strictly complementary (see Definition 2.8) and such that X_t is primal non-degenerate. Notice that the choice of factor Y_t can be arbitrary, since it does not affect any of the subsequent statements. In Lemma 4.1 and its proof, we have seen that for every X_t the unique Lagrange multiplier μ_t satisfies $Z_t = C_t - \mathcal{A}_t^*(\mu_t)$, that is,

$$\mu_t = (\mathcal{A}_t^*)^\dagger (Z_t - C_t)$$

with $(\mathcal{A}_t^*)^\dagger$ being the pseudo-inverse of $\mathcal{A}_t^*$. By assumption A1, C_t and $\mathcal{A}_t^*$ depend smoothly on t and so does $(\mathcal{A}_t^*)^\dagger$, since $\mathcal{A}_t$ is surjective for all t by assumption A2. Also, by Theorem 3.1, $t \mapsto Z_t$ is smooth. Therefore the curve $t \mapsto \mu_t$ is smooth. Since the algorithm operates in the (Y, μ) space, our implicit goal is to show that the iterates stay close to the set

$$\mathcal{S} := \{(Y_t, \mu_t) \mid (Y_t Y_t^T, Z(\mu_t)) \text{ is an optimal primal-dual pair to } (\mathrm{P}_t, \mathrm{D}_t),\ t \in [0, \tau]\}$$

containing the optimal primal-dual trajectories in the Burer-Monteiro factorization.

Lemma 4.2. *The set $\mathcal{S}$ is compact.*

Proof. As the curve $t \mapsto \mu_t$ is continuous, it suffices to prove that the set $\mathcal{S}_Y = \{Y_t \mid t \in [0, \tau]\}$ is compact. Since $\|Y_t\|_F = \sqrt{\mathrm{trace}(X_t)}$ and $t \mapsto X_t$ is smooth, it is bounded. To see that the set is closed, let $(Y_n) \subset \mathcal{S}_Y$ be a convergent sequence with limit Y such that $Y_n Y_n^T = X_{t_n}$ for some $t_n \in [0, \tau]$. By passing to a subsequence, we can assume $t_n \to t \in [0, \tau]$. Then obviously $X_t = YY^T$, which shows that Y is in the set. □

We consider the norm on $\mathbb{R}^{n\times r} \times \mathbb{R}^m \times \mathbb{S}^r_{skew}$ defined by $\|(Y, \mu, \nu)\|^2 = \|Y\|_F^2 + \|\mu\|^2 + \|\nu\|_F^2$. The induced operator norm is denoted $\|\cdot\|_{op}$.

Lemma 4.3. *There exists a constant $m > 0$ such that*

$$\|\mathcal{J}_{Y_t,t}(Y_t, \mu_t, 0)^{-1}\|_{op} \leq \frac{1}{m} \tag{4.18}$$

for all $(Y_t, \mu_t) \in \mathcal{C}$.

Proof. On its open domain of definition, the map $(Y, \mu) \mapsto \|\mathcal{J}(Y, \mu, 0)^{-1}\|_{op}$ is continuous. By Theorem 4.2, the compact set $\mathcal{C}$ is contained in that domain. Therefore, $\|\mathcal{J}(Y, \mu, 0)^{-1}\|_{op}$ achieves its maximum on $\mathcal{C}$. □

Lemma 4.4. *For any $t \in [0, \tau]$ and $\hat{Y} \in \mathbb{R}^{n\times r}$, the mapping $(Y, \mu, \nu) \mapsto \mathcal{J}_{\hat{Y},t}(Y, \mu, \nu)$ is Lipschitz continuous in operator norm on $\mathbb{R}^{n\times r} \times \mathbb{R}^m \times \mathbb{S}^r_{skew}$. Specifically,*

$$\|\mathcal{J}_{\hat{Y},t}(Y_1, \mu_1, \nu_1) - \mathcal{J}_{\hat{Y},t}(Y_2, \mu_2, \nu_2)\|_{op} \leq 12\sqrt{3}\|\mathcal{A}_t\|\|(Y_1, \mu_1, \nu_1) - (Y_2, \mu_2, \nu_2)\|$$

for all (Y_1, μ_1, ν_1) and (Y_2, μ_2, ν_2), where $\|\mathcal{A}_t\|$ is the operator norm of $\mathcal{A}_t$.

Proof. It follows from (4.17) that as a bilinear form one has

$$
\begin{aligned}
&(\mathcal{J}_{\hat{Y},t}(Y_1,\mu_1,\nu_1)-\mathcal{J}_{\hat{Y},t}(Y_2,\mu_2,\nu_2))[(H,\Delta\mu,\Delta\nu),(H,\Delta\mu,\Delta\nu)]\\
&=\operatorname{trace}(H^T\mathcal{A}_t^*(\mu_2-\mu_1)H)-2(\Delta\mu)^T\mathcal{A}_t((Y_1-Y_2)H^T+H(Y_1-Y_2)^T)\\
&\leq \|\mathcal{A}_t\|\|\mu_1-\mu_2\|\|H\|_F^2+4\|\mathcal{A}_t\|\|Y_1-Y_2\|_F\|H\|_F\|\Delta\mu\|\\
&\leq (\|\mathcal{A}_t\|\|\mu_1-\mu_2\|+4\|\mathcal{A}_t\|\|Y_1-Y_2\|_F)(\|H\|_F+\|\Delta\mu\|+\|\Delta\nu\|_F)^2\\
&\leq 4\|\mathcal{A}_t\|(\|Y_1-Y_2\|_F+\|\mu_1-\mu_2\|+\|\nu_1-\nu_2\|)(\|H\|_F+\|\Delta\mu\|+\|\Delta\nu\|_F)^2\\
&\leq 12\sqrt{3}\|\mathcal{A}_t\|\|(Y_1,\mu_1,\nu_1)-(Y_2,\mu_2,\nu_2)\|\|(H,\Delta\mu,\Delta\nu)\|^2.
\end{aligned}
$$

This proves the claim. □

Since $t\mapsto\mathcal{A}_t$ is assumed to be continuous, the constant $M=\max_{t\in[0,\tau]}12\sqrt{3}\|\mathcal{A}_t\|$ satisfies the uniform Lipschitz condition

$$
\|\mathcal{J}_{\hat{Y},t}(Y_1,\mu_1,\nu_1)-\mathcal{J}_{\hat{Y},t}(Y_2,\mu_2,\nu_2)\|_{op}\leq M\|(Y_1,\mu_1,\nu_1)-(Y_2,\mu_2,\nu_2)\| \tag{4.19}
$$

for all (Y_1,μ_1,ν_1) and (Y_2,μ_2,ν_2), independent of the choice of $\hat{Y}\in\mathbb{R}^{n\times r}$. In what follows, we proceed with using (4.19) and (4.18), without further investigating the sharpest possible bounds.

In addition, let $\lambda_r(X_t)\geq\lambda_*>0$ be a uniform lower bound on the smallest positive eigenvalue as in (4.11). Furthermore, we now also assume a uniform upper bound

$$
\|Y_t\|_2=\sqrt{\lambda_1(X_t)}\leq\sqrt{\Lambda_*}.
$$

on the spectral norm of Y_t. Finally, let $\|\dot{X}_t\|_F\leq L$ as in (4.10) and since the curve $t\mapsto\mu_t$ is smooth, the constant

$$
K:=\max_{t\in[0,\tau]}\|\dot{\mu}_t\| \tag{4.20}
$$

is also well-defined.

With the necessary constants at hand, we are now in the position to state our main result on the error analysis. The following theorem shows that we can bound the distance between the iterates of Algorithm 1 and the set of solutions to (Q_t) provided the initial point is close enough to the set of initial solutions and the stepsize Δt is small enough. Here we employ again the natural distance measure $\min_{Q\in\mathcal{O}_r}\|\hat{Y}-YQ\|_F$ between the orbits $\hat{Y}\mathcal{O}_r$ and $Y\mathcal{O}_r$, cf. Remark 9.

Theorem 4.3. *Let $\delta > 0$ and $\Delta t > 0$ be small enough such that the following three conditions are satisfied:*

$$(2\sqrt{\Lambda_*}+\delta)\delta + L\Delta t < \frac{2\lambda_*}{\sqrt{r+4}+\sqrt{r}}, \tag{4.21}$$

$$\delta < \frac{2}{3}\frac{m}{M}, \tag{4.22}$$

$$\left[\frac{1}{\lambda_*}((2\sqrt{\Lambda_*}+\delta)\delta + L\Delta t)^2 + \sqrt{r}(2\sqrt{\Lambda_*}+\delta)\delta + L\Delta t\right]^2 + (\delta + K\Delta t)^2 \leq \frac{2}{3}\frac{m}{M}\delta. \tag{4.23}$$

Assume for the initial point $(\hat{Y}_0, \hat{\mu}_0)$ that

$$\min_{Q\in\mathcal{O}_r} \|(\hat{Y}_0, \hat{\mu}_0) - (Y_0 Q, \mu_0)\| \leq \delta. \tag{4.24}$$

Then Algorithm 1 is well-defined and for all $t_{k+1} = t_k + \Delta t$ the iterates satisfy

$$\min_{Q\in\mathcal{O}_r} \|(\hat{Y}_k, \hat{\mu}_k) - (Y_{t_k} Q, \mu_{t_k})\| \leq \delta.$$

It then holds that

$$\|\hat{X}_k - X_{t_k}\|_F \leq (2\sqrt{\Lambda_*}+\delta)\delta$$

for all t_k.

Notice that the left side of (4.23) is $O(\delta^2 + \Delta t^2)$ for $\delta, \Delta t \to 0$, whereas the right side is only $O(\delta)$. Therefore for δ and Δt small enough, (4.23) will be satisfied. Furthermore, a sufficient condition for (4.24) to hold is that

$$\|\hat{\mu}_0 - \mu_0\| \leq \frac{\delta}{\sqrt{2}}$$

and

$$\|\hat{X}_0 - X_0\|_F \leq \frac{\sqrt{r\lambda_*^2 + 2\sqrt{2}\delta\lambda_*} - \sqrt{r\lambda_*^2}}{2},$$

which easily follows from (4.9).

Proof. We will investigate one step of the algorithm and apply an induction hypothesis that at time point $t = t_k$ there exists $(Y_t, \mu_t) \in \mathcal{C}$ satisfying

$$\|(\hat{Y}_t, \hat{\mu}_t) - (Y_t, \mu_t)\|_F \leq \delta.$$

We aim to show that for sufficiently small $\delta > 0$ and $\Delta t > 0$ the next iterate $(\hat{Y}_{t+\Delta t}, \hat{\mu}_{t+\Delta t})$

in the algorithm is well-defined and satisfies the same estimate

$$\|(\hat{Y}_{t+\Delta t}, \hat{\mu}_{t+\Delta t}) - (Y_{t+\Delta t}, \mu_{t+\Delta t})\|_F \le \delta$$

with an exact solution $(Y_{t+\Delta t}, \mu_{t+\Delta t}) \in \mathcal{C}$. The proof of the theorem then follows by induction over the steps in the algorithm.

We first claim that there exists an exact solution $Y_{t+\Delta t}$ in the horizontal space of $\hat{Y}_t$, that is, $X_{t+\Delta t} = Y_{t+\Delta t} Y_{t+\Delta t}^T$ and $h_{\hat{Y}_t}(Y_{t+\Delta t}) = 0$. Indeed, using (4.21) we have

$$\begin{aligned}\|\hat{X}_t - X_t\|_F &= \|(\hat{Y}_t - Y_t)\hat{Y}_t^T + Y_t(\hat{Y}_t - Y_t)^T\|_F \\ &\le (\|Y_t\|_2 + \|\hat{Y}_t\|_2)\delta \le (2\sqrt{\Lambda_*} + \delta)\delta < \frac{2\lambda_*}{\sqrt{r+4} + \sqrt{r}} - L\Delta t.\end{aligned}$$

This yields

$$\|\hat{X}_t - X_{t+\Delta t}\|_F \le \|\hat{X}_t - X_t\|_F + \|X_t - X_{t+\Delta t}\|_F < \frac{2\lambda_*}{\sqrt{r+4} + \sqrt{r}}.$$

Thus, Proposition 4.3 states the existence of $Y_{t+\Delta t}$ as desired. We note for later use that by (4.9) it satisfies

$$\begin{aligned}\|\hat{Y}_t - Y_{t+\Delta t}\|_F &\le \frac{1}{\lambda_*}\|\hat{X}_t - X_{t+\Delta t}\|_F^2 + \sqrt{r}\|\hat{X}_t - X_{t+\Delta t}\|_F \\ &\le \frac{1}{\lambda_*}(\|\hat{X}_t - X_t\|_F + L\Delta t)^2 + \sqrt{r}\|\hat{X}_t - X_t\|_F + L\Delta t \\ &\le \frac{1}{\lambda_*}((2\sqrt{\Lambda_*} + \delta)\delta + L\Delta t)^2 + \sqrt{r}(2\sqrt{\Lambda_*} + \delta)\delta + L\Delta t.\end{aligned} \tag{4.25}$$

The matrix $Y_{t+\Delta t}$ is an exact solution of ($\mathrm{Q}_{Y_t, t+\Delta t}$), and by Theorem 4.2 there is a unique Lagrange multiplier $\mu_{t+\Delta t}$ such that $\mathcal{F}_{\hat{Y}_t, t+\Delta t}(Y_{t+\Delta t}, \mu_{t+\Delta t}, 0) = 0$. By construction, the next iterate $(\hat{Y}_{t+\Delta t}, \hat{\mu}_{t+\Delta t}, \hat{\nu}_{t+\Delta t})$ in the algorithm is obtained from one step of the Newton method for solving this equation with starting point $(\hat{Y}_t, \hat{\mu}_t, 0)$. In light of (4.18) and (4.19), standard results (e.g. Theorem 1.2.5 in [102]) on the Newton method yield that under the condition

$$\|(\hat{Y}_t, \hat{\mu}_t, 0) - (Y_{t+\Delta t}, \mu_{t+\Delta t}, 0)\|_F \le \varepsilon < \frac{2}{3}\frac{m}{M}$$

one step of the method is well-defined, i.e. $\mathcal{J}_{\hat{Y}_t, t+\Delta t}(\hat{Y}_t, \hat{\mu}_t, 0)$ is invertible, and satisfies

$$\|(\hat{Y}_{t+\Delta t}, \hat{\mu}_{t+\Delta t}, \hat{\nu}_{t+\Delta t}) - (Y_{t+\Delta t}, \mu_{t+\Delta t}, 0)\|_F \le \frac{3}{2}\frac{M}{m}\|(\hat{Y}_t, \hat{\mu}_t, 0) - (Y_{t+\Delta t}, \mu_{t+\Delta t}, 0)\|_F^2.$$

In particular, using $\varepsilon = \left(\frac{2}{3}\frac{m}{M}\delta\right)^{1/2}$ would give the desired result

$$\|(\hat{Y}_{t+\Delta t}, \hat{\mu}_{t+\Delta t}) - (Y_{t+\Delta t}, \mu_{t+\Delta t})\|_F \leq \frac{3}{2}\frac{M}{m}\varepsilon^2 = \delta.$$

Therefore, we need to ensure that

$$\|(\hat{Y}_t, \hat{\mu}_t) - (Y_{t+\Delta t}, \mu_{t+\Delta t})\|_F \leq \left(\frac{2}{3}\frac{m}{M}\delta\right)^{1/2} < \frac{2}{3}\frac{m}{M}$$

is satisfied. Here the second inequality is just condition (4.22). We now show that (4.23) is a sufficient condition for the first inequality. Clearly, using (4.20), we have

$$\|\hat{\mu}_t - \mu_{t+\Delta t}\|^2 \leq (\|\hat{\mu}_t - \mu_t\| + K\Delta t)^2 \leq (\delta + K\Delta t)^2.$$

Together with (4.25) this gives

$$\|\hat{Y}_t - Y_{t+\Delta t}\|_F^2 + \|\hat{\mu}_t - \mu_{t+\Delta t}\|^2 \leq$$
$$\leq \left[\frac{1}{\lambda_*}((2\sqrt{\Lambda_*} + \delta)\delta + L\Delta t)^2 + \sqrt{r}(2\sqrt{\Lambda_*} + \delta)\delta + L\Delta t\right]^2 + (\delta + K\Delta t)^2.$$

Now (4.23) ensures the desired estimate for the right-hand and the proof is completed. □

Scene 3 Numerical experiments on Max-Cut relaxations

Where the TV-SDP princes, together with the wizard of Burer–Monteiro, arrive in a remote village harassed by the raids of the terrible Max-Cut monster, who mercilessly cuts everything it finds on its way. This infamous creature is indeed one of the scariest in the whole Optimization world. Without hesitation, our heroes show up in front of the monster's lair, holding firmly the Horizontal Spade. At the mere sight of this, the terrible beast softens. After quickly and accurately locking it in a cage, the three take it to the kingdom of Conic Optimization, where they are welcomed with a sumptuous banquet.

In this section, we address a time-varying version of the Max-Cut problem, and compare the tracking of the trajectory of solutions to TV-SDP via Algorithm 1 with interior-point methods (IPMs) used to track the same trajectory by solving the problem at discrete time points. In our experiments, we used the implementation of the homogeneous and self-dual algorithm [87, 88] from the MOSEK Optimization Suite, version 9.3 [103].

Furthermore, in order to provide a comparison with an alternative warm-start approach, we performed numerical experiments using the Splitting Conic Solver (SCS), version 3.2.2 [104]. This package implements the first-order method presented in [105, 106], which uses an operator splitting method, the alternating directions method of multipliers, to solve the homogeneous self-dual embedding. We show that the algorithm proposed can perform better, in terms of both accuracy and runtime, than repeated runs of IPM for time-invariant SDP and than the warm-started SCS.

Given a weighted graph $\mathcal{G} = (V, E)$, the Max-Cut problem is a well-known problem in graph theory. There, we wish to find a binary partition of the vertices in V (also known as a cut) of maximal weight. The weight of the cut is defined as the sum of the weights of the edges in E connecting the two subsets of the partition. This problem can be formulated as the following quadratically-constrained quadratic problem

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & \sum_{i,j=1}^{n} w_{i,j}(1 - x_i x_j) \\ \text{s.t.} \quad & x_i^2 = 1 \quad \text{for all } i \in \{1, \dots, n\}, \end{aligned} \tag{MC}$$

where $n = |V|$ is the number of vertices of the graph, $w_{i,j}$ is the weight of the edge connecting vertices i and j, and variable $x_i \in \{1, -1\}$ takes binary values according to the subset to which vertex i is assigned. This problem can be relaxed to an SDP of the form

$$\begin{aligned} \min_{X \in \mathbb{S}^n} \quad & \langle W, X \rangle \\ \text{s.t.} \quad & X_{i,i} = 1 \quad \text{for all } i \in \{1, \dots, n\} \\ & X \succeq 0, \end{aligned} \tag{MCR}$$

where W is the weights matrix whose entry (i, j) is given by $w_{i,j}$, see [5]. Note that the number of constraints is equal to the size of the variable matrix. Randomized approximation algorithms for (MC) exploiting the convex relaxation (MCR) deliver solutions with a performance ratio of 0.87 and are known to be the best poly-time algorithms to approximately solve (MC).

We adopt a time-varying version of (MCR) as a benchmark, where the data matrix W depends on a time parameter $t \in [0, 1] \mapsto W_t \in \mathbb{S}^n$. (We point out that this differs from the recently studied variant [107, 108] with edge insertions and deletions, which could be seen as discontinuous functions of time.)

In our experiment, W_t is obtained as a random linear perturbation of a sparse weight matrix with density 50%. Specifically,

$$W_t = W_0 + tW_1,$$

where the entries of W_0 are randomly generated with a normal distribution having mean and standard deviation $\nu, \sigma = 10$, while the entries in W_1 are chosen with a normal distribution having $\nu, \sigma = 1$. Both matrices have the same sparsity structure. We refer to such a problem as the *Time-Varying Max-Cut Relaxation* (TV-MCR), which can be thought of as a convex relaxation for a Max-Cut problem where the edges weights of a given graph change over time.

All the experiments were conducted on a personal computer with a 1,6 GHz Intel Core i5 dual-core processor with 16GB RAM, using a Python implementation of our path-following algorithm. The main goal was to illustrate the potential computational benefits of our algorithm, so we did not attempt to provide the most efficient implementation. The code[2] as well as the data and experimental results[3] are available online.

We performed experiments on 110 instances of the TV-MCR problem with $n = 100$ vertices and tracked the trajectory of solutions for $t \in [0,1]$. Among these samples, we included 10 instances of TV-MCR for which the rank of the solution is not constant, hence violating our assumption A5. This was done by sampling the rank (estimated with a tolerance on zero eigenvalues of 10^{-7}) of the solutions obtained using MOSEK over a 10-steps subdivision of the interval $[0,1]$ and selecting ten cases in which we observed a change in the rank. Using the same procedure, we checked that for the remaining 100 instances, the rank of the solution is constant along the trajectory.

First, we applied Algorithm 1 without stepsize adjustment, hence setting the parameter `STEPSIZE_TUNING` to `FALSE`, and using stepsizes $\Delta t = 0.1, 0.01, 0.001$, so that in each experiment 10, 100, and 1000 iterations are performed for each choice of the stepsize (see Figures 4.2 and 4.3). The factor dimension r is chosen equal to the rank of an initial solution obtained using MOSEK with relative gap termination tolerances set to 10^{-14}. Its distribution is shown in Table 4.1.

r	4	5	6	7
# occurences	2	39	53	6

Table 4.1: Distribution of the rank over 100 instances of the TV-MCR with $n = 100$ with constant rank solution trajectory.

[2] available at `github.com/antoniobellon/burer-monteiro-path-following`
[3] available at `zenodo.org/record/7769225`

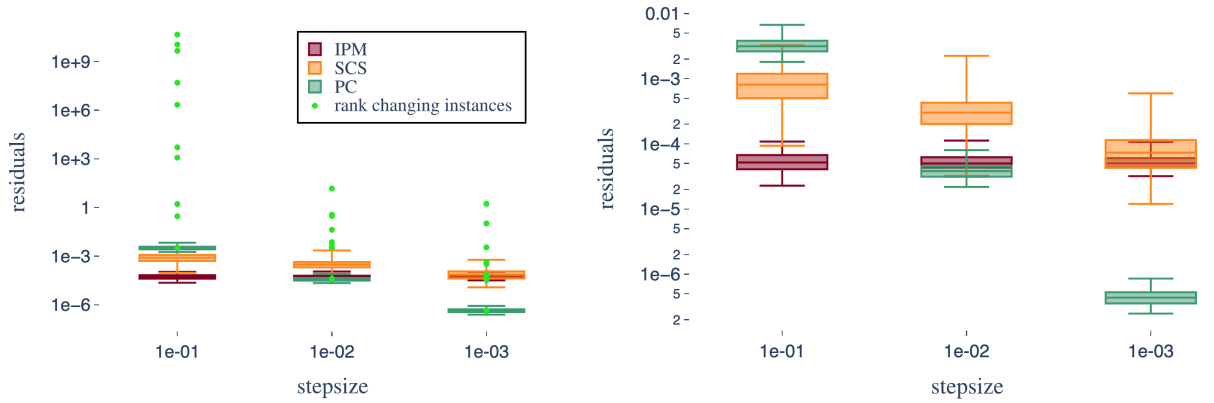


Figure 4.2: Distribution of the average residuals as a function of the stepsize.

Figure 4.2 depicts the distribution over 100 instances of the average residuals along the tracking of the solution on the time interval $[0,1]$, as a function of the used stepsize. For each whisker plot, the error bars span the interval from the minimum to the maximum, while the box spans the first quartile to the third quartile, with a horizontal line at the median.

In the left plot, the light green dots correspond to the average residuals of the 10 rank-changing instances; instead, the right plot excludes these degenerate instances form the data set. Notice that these points correspond to TV-SDP instances that do not satisfy our assumption A5. The green plot shows the average residual obtained by tracking the solution with Algorithm 1, the orange plot shows the average residual when the tracking is done using SCS with relative and absolute feasibility tolerances set to 10^{-7}, warm-started with the current solution; finally, the bordeaux color plot shows the average residual when the tracking is done using MOSEK IPM [103] with the relative gap termination tolerances set to 10^{-15}.

In analogy to (RES), the residual of an SDP primal-dual solution $(X, Z(\mu))$ is defined as

$$\mathrm{res}_t(X,\mu) := \left\| \begin{matrix} 2[C_t - \mathcal{A}_t^*(\mu)]X \\ \mathcal{A}_t(X) - b_t \end{matrix} \right\|_\infty .$$

By choosing a suitable stepsize (in our experiments order 10^{-2}), Algorithm 1 yields an average residual accuracy that is comparable to the one obtained using standard IPMs with very small relative gap termination tolerance. For stepsize of order 10^{-3}, our algorithm exhibits a residual precision that is 100 times more accurate than both IPM and warm-started SCS. Furthermore, as we see next, this accuracy is reached much faster with our approach.

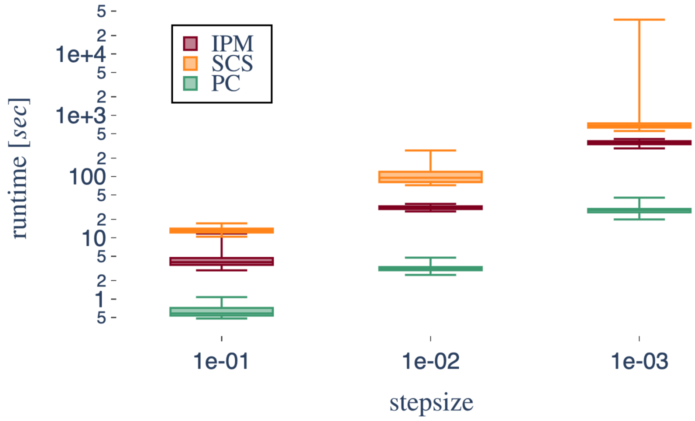


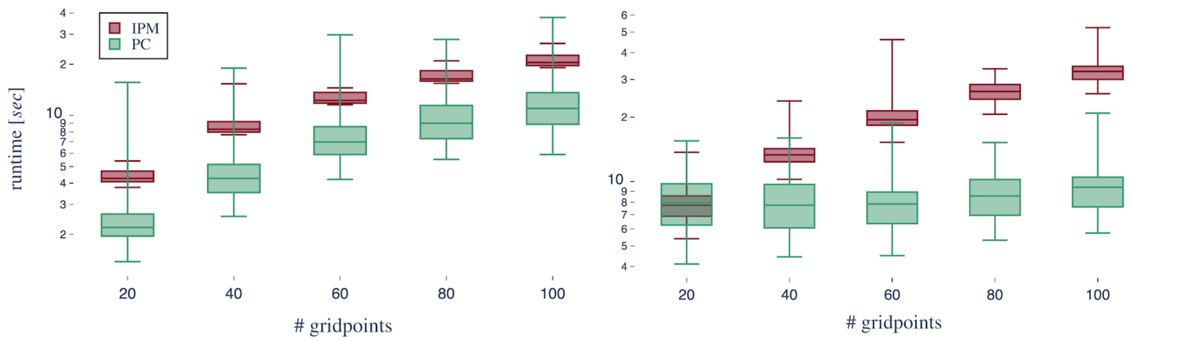
Figure 4.3: Distribution of the runtime as function of the stepsize.

In Figure 4.3 we plot the distributions of the runtimes of Algorithm 1 (green) as function of the stepsize, as well as the distributions of the runtimes of IPM (bordeaux) used with relative gap termination tolerances 10^{-15} and of the warm-started SCS (orange) to track the solutions trajectory at a constant stepsize resolution.

Remarkably, for each stepsize that we tested, the mean runtime of Algorithm 1 is on average about ten times smaller then both SCS and MOSEK IPM, indicating competitive computational performances of our algorithm.

Finally, we apply Algorithm 1 to the same set of TV-MCR problems allowing for a stepsize adjustment (setting `STEPSIZE_TUNING` to `TRUE`). In order to provide a fair comparison with MOSEK IPM, we fixed five subdivisions of the interval $[0, 1]$ in a grid of, respectively, 20, 40, 60, 80, and 100 equidistant points. For each grid, at each time point, we used MOSEK with a relative gap termination tolerance of 10^{-14} to obtain the corresponding TV-SDP solution, recording the runtime and the average residual over the tracking of each instance. For each grid, we then run our algorithm with stepsize adjustment in order to ensure the same average residual accuracy guaranteed by MOSEK, additionally enforcing the path-following procedure to hit the grid points. In this way, we ensure that our procedure has the same accuracy of MOSEK both in terms of the solution residual and of the tracking resolution.

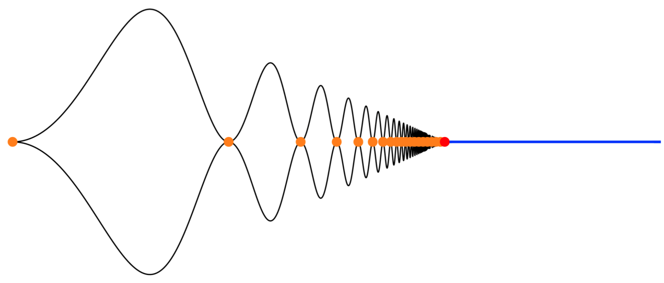
In the next page, Figure 4.4 shows the distributions of the runtimes as function of the number of grid points of both Algorithm 1 (green) and IPM with two different relative gap termination tolerances: 10^{-9} (Figure 4.4a) and 10^{-15} (Figure 4.4b).



(a) Relative gap termination tolerance $= 10^{-9}$ (b) Relative gap termination tolerance $= 10^{-15}$

Figure 4.4: Average runtime of MOSEK IPM and Algorithm 1 for tracking the TV-SDP solutions with the same residual accuracy on a grid, as a function of the number of gridpoints.

Encouragingly, we observe that we can ensure both the same accuracy and tracking resolution of MOSEK at a smaller average runtime. The constant behavior of the green plot on the right is due to the fact that, in order to ensure the same residual accuracy of the IPM, the path-following procedure needs to consider a number of points that is quite denser then the number of grid points, and hence independent from this latter, while for the plot on the left it is instead sufficient for Algorithm 1 to follow the grid.



CHAPTER 5

Coda

"Dunque è proprio finita?"

Rodolfo in La Bohème (Act III)

A discussion on the theory

Our work draws upon a long history of work in parametric optimization. In particular, the pioneering work of [63, Chapter 2] outlined a classification of solutions to univariate parametric non-linear constrained optimization problems. There, precise algebraic conditions are shown for points satisfying first-order optimality conditions to be *non-degenerate* (see Remark 8). These points exhibit a regular behavior. For *degenerate* points, four different types are defined according to which subset of non-degeneracy conditions is violated. Analogously, our approach also starts by considering algebraic conditions that ensure a regular behavior, but our classification of irregular points was made according to the local behavior of the trajectory of solutions at the point considered, rather than according to different sets of algebraic conditions.

We notice that regular points and discontinuous isolated multiple points, defined as in Definitions 3.3 and 3.5 respectively, were first identified by [58] (see e.g. Example 1 there) within the optimal partition approach to parametric analysis for linearly parametrized SDP. Non-differentiable points (Definition 3.4) can be also easily derived from their results.

Our work can hence be seen as a completion of the effort of [58]. Likewise, in our analysis, Theorem 3.4 relies on Theorem 3.2 and Theorem 3.1. There, the proof of Theorem 3.2 uses the technique of [58], while Theorem 3.1 is essentially an application of the implicit function theorem, implying that this can be applied almost everywhere. Theorem 3.3 suggests that when, instead, the assumptions for implicit function theorem do not hold almost everywhere, this allows for a broader range of possible behaviors, listed in the last row of Table 5.1.

Problem assumptions	**Type of points**
TV-SDP with LICQ, polynomial data, strict feasibility, and a generic *non-singular* time	Regular points Non-differentiable points Discontinuous isolated multiple points
TV-SDP with LICQ, continuous data, strict feasibility, without a generic *non-singular* time	Regular points Non-differentiable points Discontinuous isolated multiple points Discontinuous non-isolated multiple points Continuous bifurcation points Irregular accumulation points

Table 5.1: Assumptions on TV-SDP and associated possible type of points

From the point of view of formulating a TV-SDP, the key insight of [58] and ours is that even seemingly strong and standard assumptions such as the continuity of the data and primal-dual strict feasibility are not sufficient to prevent pathological behavior. We presented a complete characterization of such behaviors. Thereby, we showed that guaranteeing the existence of a generic non-singular point along the trajectory suffices to prevent highly pathological behaviors. However, this does not prevent from a finite number of losses of differentiability or isolated losses of uniqueness to occur.

One may also be interested in understanding how the main result of this paper specializes to restricted classes of TV-SDP, such as time-varying linear programming (TV-LP) and time-varying second order cone programming (TV-SOCP) (see Table 2.1). In the first case, if the data are assumed to be continuous functions, one can easily construct an example of each type of behaviors of the trajectory of solutions described in Definitions (3.3-3.8). For example, for $t \in (-1,1)$ consider:

1. $\min\{x \mid x \geq 1+t\}$.
2. $\min\{x \mid x \geq |t|\}$.
3. $\min\{tx \mid -1 \leq x \leq 1\}$.
4. $\min\{f(t)x \mid -1 \leq x \leq 1\}$, with $f(t) = t$ if $t \leq 0$, otherwise $f(t) = 0$.
5. $\min\{0 \cdot x \mid -g(t) \leq x \leq g(t)\}$, with $g(t) = 0$ if $t \leq 0$, otherwise $g(t) = t$.
6. $\min\{0 \cdot x \mid -h(t) \leq x \leq h(t)\}$, with $h(t) = 0$ if $t > 0$, otherwise $h(t) = t\sin^2 \frac{\pi}{t}$.

At $\hat{t} = 0$, $x^\star = 0$ is 1. a regular point, 2. a non-differentiable point, 3. an isolated discontinuous multiple point, 4. a non-isolated discontinuous multiple point, 5 a continuous bifurcation point, 6 an irregular accumulation point. Hence, restricting to the class of TV-LP does not exclude any type of point. It follows that also in the case of TV-SOCP, a class that generalize TV-LP, all the type of points can possibly appear. From this point of view, it is surprising that the trajectories of solution to TV-SDP, a class of optimization problems much wider than TV-LP, does not present, in the general framework that we adopted, any behavior which does not already show up in TV-LP. However, we believe that under a set of assumptions more specific than the one that we adopted in Theorem 3.3, some type of behaviors may be ruled out in TV-LP, but not in TV-SDP. Take as an example non-differentiable points (see Def. 3.4). If one assumes that the time dependence of the data is smooth, e.g. polynomial, non differentiable points can still appear in TV-SDP (see the first example in the last section of Chapter 3). This is due to the facial geometry of SDP, where positively curved surfaces appear, which must then entirely consist of extreme points. Instead, in TV-LP, extreme points are always isolated, so that when the solution is unique, this must coincide with a fixed extreme point. If the time dependence is smooth, this should imply that the feasible set, hence its extreme points, should also move smoothly, preventing non-differentiable points to show up. The investigation of such distinctions between TV-LP and TV-SDP may be an interesting direction for future research.

It is also worth mentioning that TV-SDP can be linked to TV-POP in a straightforward way through the Moment-Sum-of-Squares hierarchy: indeed one can substitute a TV-POP with a TV-SDP relaxation of some order, that could possibly change along the time parameterization. We conducted preliminary research to understand whether our classification results could yield non-trivial information on the original TV-POP. Unfortunately, such a line of research has not be fruitful, yet.

We used set-valued analysis to describe and study the trajectory of solutions to TV-SDP. The analysis we carried out brought us to define six different types of points, according to the local structure of the solutions trajectory. Our main result consists in proving that under standard assumptions, there are no other types of points.

One could extend our research by weakening our assumptions: continuity of the data dependence on the parameter, LICQ, and primal and dual strict feasibility throughout the parameterization interval. These requirements avoid highly degenerate situations. In particular, without continuity of the data, one can expect the trajectory to potentially present a lot of irregularities, e.g., it may fail to be both inner and outer semi-continuous, while, as Theorem 2.3 shows, under the continuity of the data outer semi-continuity is ensured. When strict feasibility is lost, two additional forms of degenerate behavior might occur: the optimal value may not be attained at any feasible point, or there may be a strictly positive duality gap between the primal and dual optimal values. It is not yet clear whether there could be other types too.

A discussion on the algorithm

In this thesis, we also proposed an algorithm for solving time-varying SDPs based on a path-following scheme for the Burer–Monteiro factorization. The restriction to a horizontal space ensures that the linearized KKT conditions system is uniquely solvable under standard regularity assumptions on the TV-SDP problem, thus leading to a well-defined path-following procedure with rigorous error bounds on the distance from the optimal trajectory. Preliminary numerical experiments on a time-varying version of the Max-Cut SDP relaxation suggest that our algorithm is competitive both in terms of runtime and accuracy when compared to the application of standard interior point methods. Future work should explore the applicability and relative merits of our approach in further applications. We would like to highlight two interesting directions for further research.

In Chapter 4, we have assumed that the rank r of the true solution curve is known and remains constant. While this is certainly appropriate for a rigorous analysis as conducted in this work, it might be restrictive in practice, as the analysis carried out in Chapter 3suggests. An important extension hence would be to develop rank-adaptive versions of our path-following approach that are able to detect and adjust the appropriate rank in a Burer–Monteiro factorization, for example by monitoring the smallest singular values of the current matrix iterate Y_t.

Another important aspect is the initialization of the method, which requires an accurate SDP solution and is currently not based on Burer–Monteiro factorization, thus undermining the computational efficiency of the whole approach. The obvious way out is to also solve the initial time problem using the factorized approach [65]. The meta-algorithm presented in [68] even does this in a rank-adaptive way. Although this is a non-convex problem, several works, including also [109, 110, 111], have been dedicated to obtaining certificates for global optimality under mild conditions, making this a reliable approach in practice.

Epilogue

Sadly, our play is almost over. All that remains is to narrate the final fate of the characters. After an outstanding academic career, the primal prince is awarded an important professorship at the Royal Conic Institute of Cartography. Years after their first encounter, the dual prince meets the Degenerate Witch, this time not in disguise. He falls in love with her even more madly than the first time, decides to marry her, and moves with her to the Singular Wood. As for the Max-Cut monster, the merciful Queen of Cones decides to release it on condition that it is willing to lend his services as a court tailor. Accepting the deal, the monster becomes a renown fashion designer for cones, ending up launching its own haute-couture brand with great success. And they lived optimally ever after!

Bibliography

[1] L. Lovász, “Semidefinite programs and combinatorial optimization,” in *Recent advances in algorithms and combinatorics*, ser. CMS Books Math./Ouvrages Math. SMC, vol. 11, Springer, 2003, pp. 137–194 (Cited on page 2).

[2] C. Helmberg, *Semidefinite programming for combinatorial optimization*. PhD thesis, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2000 (Cited on page 2).

[3] A. Majumdar, G. Hall, and A. A. Ahmadi, “Recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 331–360, 2020 (Cited on page 2).

[4] Y. Sekiguchi and H. Waki, “Perturbation analysis of singular semidefinite programs and its applications to control problems,” *Journal of Optimization Theory and Applications*, vol. 188, no. 1, pp. 52–72, 2021 (Cited on pages 2, 7, 27).

[5] M. X. Goemans and D. P. Williamson, “Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming,” *Journal of the Association for Computing Machinery*, vol. 42, no. 6, pp. 1115–1145, 1995 (Cited on pages 2, 80).

[6] D. Karger, R. Motwani, and M. Sudan, “Approximate graph coloring by semidefinite programming,” *Journal of the Association for Computing Machinery*, vol. 45, no. 2, pp. 246–265, 1998 (Cited on page 2).

[7] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*. SIAM, 1994 (Cited on page 2).

[8] S. J. Wright, *Primal-dual interior-point methods*. SIAM, 1997 (Cited on page 2).

[9] F. Jarre, “Interior-point methods for convex programming,” *Applied Mathematics and Optimization*, vol. 26, no. 3, pp. 287–311, 1992 (Cited on page 2).

[10] H. Jiang, T. Kathuria, Y. T. Lee, S. Padmanabhan, and Z. Song, “A faster interior point method for semidefinite programming,” in *2020 IEEE 61st Annual Symposium on Foundations of Computer Science*, 2020, pp. 910–918 (Cited on pages 2, 55).

[11] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of semidefinite programming: theory, algorithms, and applications*. Springer Science & Business Media, 2012, vol. 27 (Cited on pages 2, 6).

[12] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 796–817, 2001 (Cited on page 2).

[13] J. B. Lasserre, *Moments, positive polynomials and their applications*. Imperial College Press, 2010 (Cited on page 2).

[14] D. Henrion, M. Korda, and J. B. Lasserre, *The moment-SOS hierarchy. Lectures in probability, statistics, computational geometry, control and nonlinear PDEs*. World Scientific Publishing, 2020 (Cited on page 2).

[15] M. F. Anjos and J. B. Lasserre, *Handbook on semidefinite, conic and polynomial optimization*. Springer, 2011 (Cited on page 2).

[16] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine learning research*, vol. 5, no. Jan, pp. 27–72, 2004 (Cited on page 2).

[17] S. Aaronson, X. Chen, E. Hazan, S. Kale, and A. Nayak, "Online learning of quantum states," *Journal of Statistical Mechanics: Theory and Experiment*, no. 12, p. 124 019, 2019 (Cited on pages 2, 4, 5).

[18] D. Simmons-Duffin, "A semidefinite program solver for the conformal bootstrap," *Journal of High Energy Physics*, vol. 2015, no. 6, pp. 1–31, 2015 (Cited on page 2).

[19] D. Henrion, F. Kirschner, E. De Klerk, M. Korda, J.-B. Lasserre, and V. Magron, "Revisiting semidefinite programming approaches to options pricing: Complexity and computational perspectives," *INFORMS Journal on Computing*, vol. 35, no. 2, pp. 335–349, 2023 (Cited on page 2).

[20] M. R. Rajamani and J. B. Rawlings, "Estimation of the disturbance structure from data using semidefinite programming and optimal weighting," *Automatica*, vol. 45, no. 1, pp. 142–148, 2009 (Cited on page 2).

[21] A. M.-C. So and Y. Ye, "A semidefinite programming approach to tensegrity theory and realizability of graphs," in *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM, 2006, pp. 766–775 (Cited on page 2).

[22] A. Y. Alfakih, A. Khandani, and H. Wolkowicz, "Solving euclidean distance matrix completion problems via semidefinite programming," *Computational optimization and applications*, vol. 12, pp. 13–30, 1999 (Cited on page 2).

[23] R. A. Horn and C. R. Johnson, *Matrix analysis*, Second. Cambridge University Press, 2013 (Cited on pages 4, 21, 64, 65).

[24] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, pp. 92–107, 2011 (Cited on pages 4, 5).

[25] G. B. Dantzig, *Large-scale systems optimizations with application to energy. Technical report at Stanford's Systems Optimization Laboratory 77-3*, 1977 (Cited on page 4).

[26] E. J. Anderson, *A continuous model for job-shop scheduling*. PhD thesis, University of Cambridge, 1978 (Cited on page 4).

[27] Y. Nazarathy and G. Weiss, "Near optimal control of queueing networks over a finite time horizon," *Annals of Operations Research*, vol. 170, pp. 233–249, 2009 (Cited on page 4).

[28] F. Teren, "Minimum time acceleration of aircraft turbofan engines by using an algorithm based on nonlinear programming," in *NASA Technical Memorandum TM-73741, Lewis Research Center, Cleveland, Ohio*, 1977 (Cited on page 4).

[29] J. Carpentier, "Contribution to the economic dispatch problem," *Bulletin de la Société Française des Electriciens*, vol. 3, no. 8, pp. 431–447, 1962 (Cited on page 5).

[30] J. Liu, A. Bellon, A. Simonetto, M. Takáč, and J. Marecek, "Optimal power flow pursuit in the alternating current model," *arXiv:2211.02939*, 2022 (Cited on page 5).

[31] X. Liu, G. Zhao, J. Yao, and C. Qi, "Background subtraction based on low-rank and structured sparse decomposition," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2502–2514, 2015 (Cited on page 5).

[32] M. Fazel, *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2001 (Cited on page 5).

[33] S. Huang and H. Wolkowicz, "Low-rank matrix completion using nuclear norm minimization and facial reduction," *Journal of Global Optimization*, vol. 72, no. 1, pp. 5–26, 2018 (Cited on page 5).

[34] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004 (Cited on pages 6, 17, 20).

[35] S. Boyd and L. Vandenberghe, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, 1996 (Cited on page 6).

[36] M. J. Todd, "Semidefinite optimization," *Acta Numerica*, vol. 10, pp. 515–560, 2001 (Cited on page 6).

[37] E. De Klerk, *Aspects of semidefinite programming: interior point algorithms and selected applications*. Springer, 2006 (Cited on page 6).

[38] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton, "Complementarity and nondegeneracy in semidefinite programming," *Mathematical Programming*, vol. 77, no. 1, pp. 111–128, 1997 (Cited on pages 6, 22, 23, 70).

[39] M. Ramana and A. J. Goldman, "Some geometric results in semidefinite programming," *Journal of Global Optimization*, vol. 7, no. 1, pp. 33–50, 1995 (Cited on page 6).

[40] A. Shapiro, “First and second order analysis of nonlinear semidefinite programs,” *Mathematical programming*, vol. 77, pp. 301–320, 1997 (Cited on page 6).

[41] A. Simonetto, E. Dall’Anese, S. Paternain, G. Leus, and G. B. Giannakis, “Time-varying convex optimization: time-structured algorithms and applications,” *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2032–2048, 2020 (Cited on page 6).

[42] S. Gass and T. Saaty, “The computational algorithm for the parametric objective function,” *Naval Research Logistics Quarterly*, vol. 2, pp. 39–45, 1955 (Cited on page 6).

[43] J. E. Kelley Jr., “Parametric programming and the primal-dual algorithm,” *Operations Research*, vol. 7, pp. 327–334, 1959 (Cited on page 6).

[44] K. G. Murty, “Computational complexity of parametric linear programming,” *Mathematical Programming*, vol. 19, no. 2, pp. 213–219, 1980 (Cited on page 6).

[45] T. Gal, “Linear parametric programming – a brief survey,” in *Sensitivity, stability and parametric analysis*, North-Holland Publishing Co., Amsterdam, 1984, pp. 43–68 (Cited on page 6).

[46] I. Adler and R. D. C. Monteiro, “A geometric view of parametric linear programming,” *Algorithmica. An International Journal in Computer Science*, vol. 8, no. 2, pp. 161–176, 1992 (Cited on page 6).

[47] R. Bellman, “Bottleneck problems and dynamic programming,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 39, pp. 947–951, 1953 (Cited on page 6).

[48] D. Bertsekas, *Dynamic programming and optimal control: Volume I*. Athena scientific, 2012, vol. 4 (Cited on page 6).

[49] E. A. Yıldırım and M. Todd, “Sensitivity analysis in linear programming and semidefinite programming using interior-point methods,” *Mathematical Programming*, vol. 90, no. 2, pp. 229–261, 2001 (Cited on page 6).

[50] A. G. Hadigheh, O. Romanko, and T. Terlaky, “Sensitivity analysis in convex quadratic optimization: Simultaneous perturbation of the objective and right-hand-side vectors,” *Algorithmic Operations Research*, vol. 2, no. 2, pp. 94–111, 2007 (Cited on page 6).

[51] D. Goldfarb and K. Scheinberg, “On parametric semidefinite programming,” *Applied Numerical Mathematics*, vol. 29, no. 3, pp. 361–377, 1999 (Cited on pages 6, 7, 8, 32).

[52] A. A. Ahmadi and B. El Khadir, “Time-varying semidefinite programs,” *Mathematics of Operations Research*, vol. 46, no. 3, pp. 1054–1080, 2021 (Cited on pages 6, 7, 8, 32).

[53] R. Al-Salih and M. Bohner, "Linear programming problems on time scales," *Applicable Analysis and Discrete Mathematics*, vol. 12, no. 1, pp. 192–204, 2018 (Cited on page 6).

[54] X. Wang, S. Zhang, and D. D. Yao, "Separated continuous conic programming: Strong duality and an approximation algorithm," *SIAM Journal on Control and Optimization*, vol. 48, no. 4, pp. 2118–2138, 2009 (Cited on page 7).

[55] M. V. Nayakkankuppam and M. L. Overton, "Conditioning of semidefinite programs," *Mathematical Programming*, vol. 85, no. 3, pp. 525–540, 1999 (Cited on page 7).

[56] A. Mohammad-Nezhad, *Conic optimization: optimal partition, parametric, and stability analysis*. PhD thesis, Lehigh University, 2019 (Cited on page 7).

[57] A. Mohammad-Nezhad and T. Terlaky, "Parametric analysis of semidefinite optimization," *Optimization*, vol. 69, no. 1, pp. 187–216, 2020 (Cited on page 7).

[58] J. D. Hauenstein, A. Mohammad-Nezhad, T. Tang, and T. Terlaky, "On computing the nonlinearity interval in parametric semidefinite optimization," *Mathematics of Operations Research*, vol. 47, no. 4, pp. 2989–3009, 2022 (Cited on pages 7, 8, 26, 32, 35, 85, 86).

[59] B. El Khadir, *Semidefinite representations in semialgebraic optimization and dynamics-oriented learning*. PhD thesis, Princeton University, 2020 (Cited on pages 7, 8).

[60] E. A. Yildirim, "Unifying optimal partition approach to sensitivity analysis in conic optimization," *Journal of Optimization Theory and Applications*, vol. 122, no. 2, pp. 405–423, 2004 (Cited on page 8).

[61] A. Bellon, D. Henrion, V. Kungurtsev, and J. Mareček, "Parametric semidefinite programming: geometry of the trajectory of solutions," to appear in *Mathematics of Operations Research*, preprint in *arXiv:2104.05445*, 2021 (Cited on pages 8, 25, 29, 57).

[62] A. Bellon, M. Dressler, V. Kungurtsev, J. Mareček, and A. Uschmajew, "Time-varying semidefinite programming: path following a Burer–Monteiro factorization," to appear in *SIAM Journal on Optimization*, preprint in *arXiv:2210.08387*, 2022 (Cited on page 8).

[63] J. Guddat, F. Guerra Vazquez, and H. T. Jongen, *Parametric optimization: singularities, pathfollowing and jumps*. Springer, 1990 (Cited on pages 8, 10, 29, 31, 39, 40, 56, 85).

[64] E. L. Allgower and K. Georg, *Introduction to numerical continuation methods*. SIAM, 2003 (Cited on pages 10, 56).

[65] S. Burer and R. D. C. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Mathematical Programming*, vol. 95, no. 2, Ser. B, pp. 329–357, 2003 (Cited on pages 10, 11, 56, 88).

[66] S. Burer and R. D. C. Monteiro, "Local minima and convergence in low-rank semidefinite programming," *Mathematical Programming*, vol. 103, no. 3, Ser. A, pp. 427–444, 2005 (Cited on pages 10, 56, 59).

[67] I. Waldspurger and A. Waters, "Rank optimality for the Burer–Monteiro factorization," *SIAM Journal on Optimization*, vol. 30, no. 3, pp. 2577–2602, 2020 (Cited on pages 11, 56).

[68] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre, "Low-rank optimization on the cone of positive semidefinite matrices," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2327–2351, 2010 (Cited on pages 11, 56, 59, 88).

[69] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "A certifiably correct algorithm for synchronization over the special euclidean group," in *Algorithmic Foundations of Robotics XII*. Cham: Springer International Publishing, 2020, pp. 64–79 (Cited on pages 11, 56).

[70] A. Barvinok, "Problems of distance geometry and convex properties of quadratic maps," *Discrete & Computational Geometry*, vol. 13, no. 2, pp. 189–202, 1995 (Cited on pages 11, 56).

[71] G. Pataki, "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues," *Mathematics of Operations Research*, vol. 23, no. 2, pp. 339–358, 1998 (Cited on pages 11, 56).

[72] N. Boumal, V. Voroninski, and A. Bandeira, "The non-convex Burer–Monteiro approach works on smooth semidefinite programs," *Advances in Neural Information Processing Systems*, vol. 29, pp. 2765–2773, 2016 (Cited on pages 11, 56).

[73] N. Boumal, V. Voroninski, and A. S. Bandeira, "Deterministic guarantees for Burer–Monteiro factorizations of smooth semidefinite programs," *Communications on Pure and Applied Mathematics*, vol. 73, no. 3, pp. 581–608, 2020 (Cited on pages 11, 56).

[74] D. Cifuentes, "On the Burer–Monteiro method for general semidefinite programs," *Optimization Letters*, vol. 15, no. 6, pp. 2299–2309, 2021 (Cited on pages 11, 56).

[75] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*. Springer, 2009 (Cited on pages 25, 27).

[76] J. N. Franklin, *Matrix theory*. Courier Corporation, 2012 (Cited on page 26).

[77] W. W. Hogan, "Point-to-set maps in mathematical programming," *SIAM Review*, vol. 15, no. 3, pp. 591–603, 1973 (Cited on page 27).

[78] D. Goldfarb and K. Scheinberg, "Interior point trajectories in semidefinite programming," *SIAM Journal on Optimization*, vol. 8, no. 4, pp. 871–886, 1998 (Cited on page 32).

[79] J. D. Hauenstein, A. C. Liddell Jr, S. McPherson, and Y. Zhang, “Numerical algebraic geometry and semidefinite programming,” *Results in Applied Mathematics*, vol. 11, p. 100 166, 2021 (Cited on page 32).

[80] J. Nie, K. Ranestad, and B. Sturmfels, “The algebraic degree of semidefinite programming,” *Mathematical Programming*, vol. 122, pp. 379–405, 2010 (Cited on page 33).

[81] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton, “Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results,” *SIAM Journal on Optimization*, vol. 8, no. 3, pp. 746–768, 1998 (Cited on page 34).

[82] J.-P. Haeberly, “Remarks on nondegeneracy in mixed semidefinite-quadratic programming.,” *Unpublished memorandum*, 1998 (Cited on page 34).

[83] S. G. Krantz and H. R. Parks, *The implicit function theorem: history, theory, and applications*. Springer Science & Business Media, 2002 (Cited on page 34).

[84] S. Basu, R. Pollack, and M.-F. Roy, *Algorithms in real algebraic geometry*. Springer, 2006 (Cited on page 35).

[85] F. Jarre, “An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices,” *SIAM Journal on Control and Optimization*, vol. 31, no. 5, pp. 1360–1377, 1993 (Cited on page 55).

[86] L. Tunçel, “Potential reduction and primal-dual methods,” in *Handbook of semidefinite programming*, vol. 27, Kluwer Academic Publishers, 2000, pp. 235–265 (Cited on page 55).

[87] E. D. Andersen, C. Roos, and T. Terlaky, “On implementing a primal-dual interior-point method for conic quadratic optimization,” *Mathematical Programming*, vol. 95, pp. 249–277, 2003 (Cited on pages 55, 79).

[88] R. M. Freund, “On the behavior of the homogeneous self-dual model for conic convex optimization,” *Mathematical Programming*, vol. 106, no. 3, Ser. A, pp. 527–545, 2006 (Cited on pages 55, 79).

[89] J. Gondzio and A. Grothey, “Reoptimization with the primal-dual interior point method,” *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 842–864, 2002 (Cited on page 56).

[90] J. Gondzio and A. Grothey, “A new unblocking technique to warmstart interior point methods based on sensitivity analysis,” *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1184–1210, 2008 (Cited on page 56).

[91] A. Engau, M. F. Anjos, and A. Vannelli, “On interior-point warmstarts for linear and combinatorial optimization,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1828–1861, 2010 (Cited on page 56).

[92] M. Colombo, J. Gondzio, and A. Grothey, “A warm-start approach for large-scale stochastic linear programs,” *Mathematical Programming*, vol. 127, no. 2, Ser. A, pp. 371–397, 2011 (Cited on page 56).

[93] A. Skajaa, E. D. Andersen, and Y. Ye, “Warmstarting the homogeneous and self-dual interior point method for linear and conic quadratic problems,” *Mathematical Programming Computation*, vol. 5, no. 1, pp. 1–25, 2013 (Cited on page 56).

[94] E. Massart and P.-A. Absil, “Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices,” *SIAM Journal on Matrix Analysis and Applications*, vol. 41, no. 1, pp. 171–198, 2020 (Cited on pages 57, 59, 61, 65).

[95] J. Im and H. Wolkowicz, “A strengthened Barvinok–Pataki bound on SDP rank,” *Operations Research Letters*, vol. 49, no. 6, pp. 837–841, 2021 (Cited on page 58).

[96] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008 (Cited on pages 59, 60).

[97] R. Balan and C. B. Dock, “Lipschitz analysis of generalized phase retrievable matrix frames,” *SIAM Journal on Matrix Analysis and Applications*, vol. 43, no. 3, pp. 1518–1571, 2022 (Cited on pages 59, 64).

[98] B. A. Schmitt, “Perturbation bounds for matrix square roots and Pythagorean sums,” *Linear Algebra and its Applications*, vol. 174, pp. 215–227, 1992 (Cited on page 64).

[99] J. Nocedal and S. Wright, *Numerical optimization*. Springer, 2006 (Cited on page 71).

[100] S. J. Wright, “An algorithm for degenerate nonlinear programming with rapid local convergence,” *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 673–696, 2005 (Cited on page 72).

[101] Q. T. Dinh, C. Savorgnan, and M. Diehl, “Adjoint-based predictor-corrector sequential convex programming for parametric nonlinear optimization,” *SIAM Journal on Optimization*, vol. 22, no. 4, pp. 1258–1284, 2012 (Cited on page 74).

[102] Y. Nesterov, *Lectures on convex optimization (second edition)*. Springer, 2018 (Cited on page 78).

[103] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual, version 9.3*. `http://docs.mosek.com/9.3/toolbox/index.html`, 2022 (Cited on pages 79, 82).

[104] B. O’Donoghue, E. Chu, N. Parikh, and S. Boyd, *SCS: Splitting Conic Solver, version 3.2.2*, `https://github.com/cvxgrp/scs`, 2022 (Cited on page 80).

[105] B. O'Donoghue, "Operator splitting for a homogeneous embedding of the linear complementarity problem," *SIAM Journal on Optimization*, vol. 31, pp. 1999–2023, 2021 (Cited on page 80).

[106] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *Journal of Optimization Theory and Applications*, vol. 169, no. 3, pp. 1042–1068, 2016 (Cited on page 80).

[107] E. Kao, V. Gadepally, M. Hurley, M. Jones, J. Kepner, S. Mohindra, P. Monticciolo, A. Reuther, S. Samsi, W. Song, D. Staheli, and S. Smith, "Streaming graph challenge: stochastic block partition," in *2017 IEEE High Performance Extreme Computing Conference*, 2017, pp. 1–12 (Cited on page 80).

[108] M. Henzinger, A. Noe, and C. Schulz, "Practical fully dynamic minimum cut algorithms," in *2022 Proceedings of the Symposium on Algorithm Engineering and Experiments*, 2022, pp. 13–26 (Cited on page 80).

[109] N. Boumal, "A Riemannian low-rank method for optimization over semidefinite matrices with block-diagonal constraints," preprint in *arXiv:1506.00575*, 2015 (Cited on page 88).

[110] D. M. Rosen, "Scalable low-rank semidefinite programming for certifiably correct machine perception," in *Algorithmic Foundations of Robotics XIV*, Springer, 2021, pp. 551–566 (Cited on page 88).

[111] D. Cifuentes and A. Moitra, "Polynomial time guarantees for the Burer-Monteiro method," in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 23 923–23 935 (Cited on page 88).