Projection Based Methods for Conic Linear Programming

—

Optimal First Order Complexities
and
Norm Constrained Quasi Newton Methods

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Felix Lieder

aus Bremen

Düsseldorf, Juni 2018

aus dem Mathematischen Institut
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1: Prof. Dr. Florian Jarre
Heinrich-Heine-Universität Düsseldorf

2: Prof. Dr. Mirjam Dür
Universität Augsburg

3: Prof. Dr. Christian Kanzow
Universität Würzburg

Tag der mündlichen Prüfung: 31. August 2018

Abstract

This thesis is devoted to solving linear conic optimization problems via (accelerated) projection based algorithms. Making use of the existing duality theory and Moreau's theorem, we reformulate the sufficient optimality conditions via the nonexpansive gradient of a reduced Lagrangian function, whose zeros are in a one to one relationship with the optimal solutions of the initial problem. The reduced Lagrangian is nonconvex and therefore the zeros generally are saddle points. We proceed by investigating first order methods to tackle this saddle point problem: A simple unitary transformation of the gradient is firmly nonexpansive allowing the application of standard fixed point methods. The analysis of these methods leads to the first main contribution of this thesis, which is the development of a general concept for analyzing convergence rates of fixed point methods for Lipschitz continuous mappings in an optimal fashion, resulting in unimprovable complexity estimates. Specifically we derive a novel mathematical toolbox for obtaining optimal worst-case complexities, which is exemplarily used to establish the optimal worst-case complexity of the so-called Krasnoselski-Mann (KM) iteration with fixed step length in real Hilbert spaces. Furthermore we address applications in designing fixed step methods with optimized worst-case or average-case complexity as well as extensions to complex spaces. These first order methods are then complemented by second order methods, leading to the second main contribution of this thesis: The design of a norm constrained limited memory quasi Newton method, which in combination with the KM iteration resulted in a competitive software package written in MATLAB. This limited memory quasi Newton method uses a (low dimensional) semidefiniteness constraint for the correction and a least-squares approach to determine the information to be dropped from memory. We perform numerical experiments on a variety of problems, including semidefinite (SDP) and doubly non-negative (DNN) programs, which mostly arise from relaxations of NP-hard problems. First the fixed point approach is applied to about 80 large scale SDP/DNN test problems and it proves to be very competitive for these problems. Finally a combination with the novel norm constrained limited memory quasi Newton method leads to a further acceleration of our implementation.

Key words: conic programs, Krasnoselski-Mann iteration, tight worst-case complexity, norm constrained quasi Newton method

Kurzfassung

Diese Dissertation beschäftigt sich mit dem Lösen von konischen Optimierungsproblemen mit Hilfe von (beschleunigten) projektionsbasierten Algorithmen. Unter Ausnutzung der existierenden Dualitätstheorie und Moreau's Theorem, werden wir die hinreichenden Optimalitätsbedingungen durch den nicht expansiven Gradienten einer reduzierten Lagrange Funktion reformulieren, dessen Nullstellen in einer eins zu eins Verbindung mit den optimalen Lösungen des initialen Problems stehen. Da die reduzierte Lagrange Funktion im Allgemeinen nicht konvex ist, handelt es sich bei den Nullstellen um Sattelpunkte. Um dieses Sattelpunkt-Problem zu lösen, fahren wir mit der Untersuchung von Verfahren erster und zweiter Ordnung fort: Eine einfache orthogonale Transformation des Gradienten ist "firmly nonexpansive" und erlaubt die Anwendung von standard Fixpunkt Verfahren. Die Analyse dieser Vefahren führt zum ersten Hauptbeitrag dieser Arbeit, welcher in der Bereitstellung eines mathematischen Konzepts besteht um Konvergenzraten von Fixpunkt Verfahren für Lipschitz stetige Funktionen auf optimale Weise zu analysieren und zu unverbesserbaren Komplexitätschranken führt. Dieses neue mathematische Werkzeug nutzen wir exemplarisch um eine optimale "worst-case" Komplexität für die sogenannte Krasnoselski-Mann (KM) Iteration mit konstanter Schrittweite in reellen Hilbert-Räumen herleiten. Wir werden außerdem mögliche Anwendungen in Bezug auf das Herleiten neuer Verfahren mit fixierten Schrittweiten sowohl für optimiertes "worst-case-" als auch durchschnittlichem Verhalten, sowie Erweiterungen für komplexe Räume ansprechen. Diese Verfahren erster Ordnung werden dann durch Verfahren zweiter Ordnung ergänzt, welche den zweiten Hauptbeitrag dieser Arbeit darsteken: Dem Design einer neuen Klasse von Norm restringierten Quasi-Newton-Verfahren mit limitiertem Speicher, welche in Verbindung mit der KM Iteration in einem konkurrenzfähigem Software Packet für MATLAB resultiert. Dieses normbeschränkte quasi Newton Verfahren mit limitiertem Speicher benutzt eine (niedrig dimensionale) semidefinite Ungleichung für die Korrektur und einen kleinste Quadrate Ansatz um Festzustellen welche Informationen den Speicher verlassen sollen. Wir führen numerische Experimente an einer Auswahl von Problemen, einschließlich semidefiniter (SDP) und doppelt nicht negativer (DNN) Programme durch, welche meist aus Relaxierungen von NP-schweren Problemen entstehen. Zuerst benutzen wir den Fixpunkt Ansatz für circa 80 hochdimensionale SDP/DNN Testprobleme, für die sich der Ansatz als hochkompetetiv herraustellt. Schließlich führt eine Kombination mit dem neuen normbeschränkten quasi Newton Verfahren zu einer weitereren Beschleunigung unserer Implementation.

Stichwörter: Konische Programme, Krasnoselski-Mann Iteration, scharfe worst-case Komplexität, normbeschränktes quasi Newton Verfahren

Contents

1 Introduction

Let $(E, \langle .,. \rangle)$ be a Euclidean space, i.e. a real finite dimensional vector space E, equipped with the symmetric inner product $\langle .,. \rangle : E \times E \to \mathbb{R}$. Let $K \subset E$ be a nonempty closed convex cone, i.e. $\mathcal{K}$ is closed and $X, Y \in \mathcal{K}$ implies the membership $\alpha X + \beta Y \in \mathcal{K} \ \forall \alpha, \beta \geq 0$. In this thesis we analyze projection approaches for solving conic optimization problems in standard primal form

$$\begin{aligned} \underset{X \in E}{\text{minimize}} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \mathcal{A}(X) = b \\ & X \in \mathcal{K} \end{aligned} \tag{1}$$

where $C \in E$, $b \in \mathbb{R}^m$, $\mathcal{A} : E \to \mathbb{R}^m$ is a linear operator and $\mathcal{A}^* : \mathbb{R}^m \to E$ is the adjoint operator of $\mathcal{A}$ (i.e. $\mathcal{A}^*$ is the unique linear operator that satisfies $y^T \mathcal{A}(X) = \langle \mathcal{A}^*(y), X \rangle \ \forall \, y \in \mathbb{R}^m, \ X \in E$). Focussing on this specific form is due to two reasons: First the wide field of application, since generically convex, but also some problems that appear nonconvex at first glance can be equivalently stated in the form above, see [22] for a survey. Examples range from "easy" to NP-hard problems, namely linear (LP), second order (SOP) and semidefinite (SDP), but also completely positive (CPP) and copositive (COP) optimization problems. Secondly this form is appealing because of its simplicity and well established duality theory, which we are going to exploit in the following. Specifically the dual problem of (1) takes the form

$$\begin{aligned} \underset{Y \in E, \ y \in \mathbb{R}^m}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & \mathcal{A}^*(y) - Y = C \\ & Y \in \mathcal{K}^P \end{aligned} \tag{2}$$

where $\mathcal{K}^P := \{Y \in E \mid \langle Y, X \rangle \leq 0, \forall X \in \mathcal{K}\} \subset E$ denotes the polar cone. Note that the polar is used instead of the dual cone $\mathcal{K}^D := -\mathcal{K}^P$ for reasons of convenience later on. For primal feasible $X \in \mathcal{K}$ with $\mathcal{A}(X) = b$ and dual feasible $Y \in \mathcal{K}^P, \ y \in \mathbb{R}^m$ with $\mathcal{A}^*(y) - Y = C$ we immediately obtain the so-called weak duality

$$b^T y - \langle C, X \rangle = \mathcal{A}(X)^T y - \langle \mathcal{A}^*(y) - Y, X \rangle = \langle X, Y \rangle \leq 0 \tag{3}$$

implying that the optimal value of (1) will always be greater or equal than the optimal value of (2), once both problems have feasible points. From (3) we can directly conclude that points $X, Y \in E$, $y \in \mathbb{R}^m$ that satisfy the sufficient optimality conditions, the so called Karush-Kuhn-Tucker (KKT) conditions

$$\begin{aligned} & \mathcal{A}(X) = b \\ & \mathcal{A}^*(y) - Y = C \\ & X \in \mathcal{K}, \ Y \in \mathcal{K}^P, \quad \langle X, Y \rangle = 0 \end{aligned} \tag{4}$$

must be optimal solutions to (1) respectively (2). Such feasible points will be referred to as KKT-points or optimal solutions in the following. Unfortunately conditions (4) are in general not necessary optimality conditions: This usually depends on further assumptions on either the involved cone or on some regularity assumptions regarding the feasible set, e.g. Slater's condition. The problem herein is that interesting problems, for example problems in [3], while possibly not satisfying these regularity assumptions, may still possess KKT-points. For our further research we will therefore rather assume existence of points satisfying (4), than enforce any other additional and limiting assumptions. Under this existence assumption, solving (1) is equivalent to finding $X, Y \in E$, $y \in \mathbb{R}^m$ satisfying (4).

This thesis is organized as follows: We start by recalling preliminaries, such as the fundamental concept of projections in section two, which is then used to define a reduced Lagrangian. Zeros of its (firmly nonexpansive) mirrored gradient are in a one to one relationship to the optimal solutions of (1) and (2), allowing us to regard solving (4) as a fixed point problem. We end the second section by recalling the concept of generalized derivatives (which are later needed for section five). After presenting a selection of relevant cones in section three, we will consider first order methods in section four. By starting with some motivational numerical considerations and an exemplarily treatment of the (mirrored) gradient of the reduced Lagrangian, we obtain an overview about known results regarding the Krasnoselski-Mann iteration (KM iteration). We will then proceed by presenting a new (optimal/unimprovable) result regarding the convergence rate of the KM-iteration with fixed step length. Its proof leads us to a general concept of proving (optimal) worst-case convergence rates for so called Fixed-step-methods (FSMs). We then extend this concept in various directions, providing a mathematical toolbox for analyzing FSMs in an optimal fashion. In section five we shift our focus back to finding zeros of the gradient of the reduced Lagrangian, this time by means of second order methods. We briefly discuss the application of generalized Newton methods but soon switch to a (new) modification of quasi Newton methods, namely norm constrained quasi Newton methods. It should be noted that in contrast to first order, the second order methods present a rather heuristic approach, as convergence results rely usually on unreasonably strong assumptions. We will nevertheless investigate computationally affordable realizations which are motivated by usual limited memory variants, consider some further numerical results in section six and finally conclude in section seven.

At this point we like to emphasize the points that are of interest in this thesis, and points that are not. Let us therefore revisit some historical facts concerning algorithms for solving (1). In 1948 Dantzig proposed his celebrated simplex algorithm to solve linear problems[1], i.e. $\mathcal{K} = \mathbb{R}^n_+$ in our notation. Variants have been implemented in a fast and reliable fashion, see for example [15], [26] or [61], making this algorithm a powerful tool and probably one of the most applied algorithms in optimization. There are however two issues: First it is still not clear whether the simplex algorithm is a polynomial time algorithm or not, which in practice seems rather negligible, since it behaves well on most real world problems. Secondly and more significantly the simplex-algorithm does not generalize to non-polyhedral cones, which is a real limitation if one wants to solve more general optimization problems. It is due to Karmarkar, who tackled both issues with his algorithm [39] in 1984: His method, an interior point method (IPM), solves LPs in polynomial time[2] and generalizes to more

[1]Dantzig started working on his algorithm in 1947. He presented it in 1948 at the meeting of the Econometric Society at the University of Wisconsin in Madison and published it in 1949. For a historical survey see[16], for Dantzig's Publications from 1949 see [17] [18] [19].

[2]Khachian [40] had already established in 1979 that LPs could be solved via the ellipsoid method.

general convex cones. For example to the semidefinite cone $\mathcal{K} = \mathbb{S}^n_+$ where

$$\mathbb{S}^n_+ := \{X \in \mathbb{R}^{n\times n} \mid X = X^T,\ y^T X y \geq 0\ \forall y \in \mathbb{R}^n\} \tag{5}$$

within the space $E = \mathbb{S}^n$ for $\mathbb{S}^n := \{X \in \mathbb{R}^{n\times n} \mid X = X^T\}$ of real symmetric matrices equipped with the inner product $\langle X, Y\rangle = \text{trace}\,(X^T Y)$. For small and medium sized SDPs, reliable variants have been implemented, see for example [77], [86], [84], [96]. One of the drawbacks of these IPMs is the following bottleneck: At every iteration one has to compute, store and factor a Schur complement matrix of dimension $m \times m$, which requires substantial computational effort if m is large. To overcome this issue, various approaches have been proposed, including proximal-point [63]; proximal alternating direction [91], [79]; augmented Lagrangian [99], [97]; spectral bundle [28]; inexact interior point [88]; augmented Lagrangian penalty [69], [56]; quasi Newton [20]; low rank factorization [4]; augmented primal dual[32] and many more methods (this is just a rather incomplete list). Very roughly all these methods can be divided into two subcategories: methods that use second order information of the conic structure, and methods that don't. While the latter methods usually avoid a changing Schur complement matrix, they cause a new bottleneck: Frequently suffering from slow convergence near the optimal solution, variables have to be factored (or projected) more often (for example $X = VV^T$ in the semidefinite case). Which method is suited best for which problem, is then quite a hard question to answer. There are at least two unsatisfying answers, a practical and a theoretical one: From a practical point of view one would say the best method, is the one that solves all given problems to a given accuracy in the minimum amount of time. This answer is however questionable: First it relies on the given problems, so the best method for a set of problems, might not be the best method for a different set of problems. Secondly we have to be more precise about the minimum amount of time: Do we mean average time? Do we mean minimum time for each problem independently? How does given hard- and software influence timing? What about implementation details?

The theoretical answer to the question of the best method would be the following: The best method is the one with the lowest complexity, i.e. the method that involves the lowest amount of numerical floating operations to solve any problem (of fixed size) to a given accuracy. There are however two major issues with this answer: First it is usually very hard to prove universal complexity bounds, and secondly, again from a practical point of view, one is often more interested in an "average" convergence behavior (e.g. as for the simplex algorithm above). Our framework presented in section four will actually tackle both issues, finding worst-case complexities and providing a new statistical point of entry into designing first order methods with good "average" performance. We do so by essentially separating the problem of finding KKT-points (4), into two problems, the first is evaluating a certain function (the mirrored gradient of the reduced Lagrangian) and the second is finding a zero of this function via some method. Interestingly we will see that the method part is significantly easier then the evaluation part. Now instead of giving another questionable answer to the question of the best method, let us collect some desirable properties, that we would like to accomplish with our first order methods.

1. We would like some kind of convergence guarantee, at least when provided with sufficient computational accuracy.
2. The method should be simple and not demand unreasonable high computational resources, ideally a linear demand in the problems dimension.

3. We would like a low worst case complexity (but not too low, see section 4.2.1).
4. We would like our method to have good average behavior, i.e. the majority of problems should be solved significantly faster than the worst case complexity suggests.

Let us now end the philosophical discussion and start with our analysis.

2 Preliminaries

In this section we will lay the foundations of our framework by recalling the well known concept of projections onto convex sets, introducing the reduced Lagrangian, investigating some of its properties and also by recalling the concept of generalized derivatives.

2.1 Projections and the Generalized Absolute Value

Let us start with the most fundamental concept needed for our further analysis: For a nonempty closed convex set $\mathcal{C} \subset E$ we define the orthogonal projection $\Pi_{\mathcal{C}}(X)$ of an arbitrary point $X \in E$ onto $\mathcal{C}$ as

$$\Pi_{\mathcal{C}}(X) := \operatorname{argmin}\{\|X - Y\| \ \mid Y \in \mathcal{C}\} \tag{6}$$

where $\|.\| := \sqrt{\langle ., . \rangle}$ denotes the induced norm. It is well know that (6) is well defined and possesses refined and well studied properties. Some of these results are collected in the following.

Lemma 2.1 (Properties of convex Projections)**.** *Let $\mathcal{C} \subset E$ be a nonempty closed convex set. The orthogonal projection $\Pi_{\mathcal{C}} : E \to E$*

- *can be characterized as follows: for $Z \in E$ and $X \in \mathcal{C}$ we have $X = \Pi_{\mathcal{C}}(Z)$ if and only if $\langle Z - X, Y - X \rangle \leq 0 \quad \forall Y \in \mathcal{C}$*
- *is firmly nonexpansive (1-cocoercive), i.e. satisfies*

$$\|\Pi_{\mathcal{C}}(X) - \Pi_{\mathcal{C}}(Y)\|^2 \leq \langle \Pi_{\mathcal{C}}(X) - \Pi_{\mathcal{C}}(Y), X - Y \rangle \quad \forall X, Y \in E$$

- *is nonexpansive (Lipschitz continuous with modulo one), i.e. satisfies*

$$\|\Pi_{\mathcal{C}}(X) - \Pi_{\mathcal{C}}(Y)\| \leq \|X - Y\| \ \ \forall X, Y \in E$$

- *positively homogeneous if $\mathcal{C} = \mathcal{K}$ is also a cone, i.e. $\Pi_{\mathcal{K}}(\lambda Z) = \lambda \Pi_{\mathcal{K}}(Z) \quad \forall \lambda \geq 0$*

The function $Z \mapsto \frac{1}{2}\|Z - \Pi_{\mathcal{C}}(Z)\|^2$ is convex, differentiable and has the derivative $Z \mapsto Z - \Pi_{\mathcal{C}}(Z)$.

Proof. As was already mentioned these results are widely known. For a proof see for example [98] Lemma 1.1 (characterization), Lemma 1.2 (cocoercivity), Theorem 4.1 (differentiability). Nonexpansiveness follows from cocoercivity in combination with the Cauchy-Schwarz inequality. ∎

In section three we will consider a selected collection of convex cones as well as their projections. Let us consider two (more fundamental) explicit examples here, namely weighted projections onto an affine subspaces. For both examples it is convenient to introduce the Moore-Penrose pseudo-inverse $\mathcal{M}^+$ of a linear operator $\mathcal{M}$, i.e. $\mathcal{M}^+$ denotes the unique linear operator that satisfies $\mathcal{M}\mathcal{M}^+\mathcal{M} = \mathcal{M}$, $\mathcal{M}^+\mathcal{M}\mathcal{M}^+ = \mathcal{M}^+, (\mathcal{M}\mathcal{M}^+)^* = \mathcal{M}\mathcal{M}^+$ and $(\mathcal{M}^+\mathcal{M})^* = \mathcal{M}^+\mathcal{M}$.

Example 2.2 (Primal Weighted Projection onto an affine Subspace)**.** *Let $\mathcal{W} : E \to E$ be some self-adjoint invertible Linear Operator. Assume that the affine space $\mathcal{L}_b := \{X \in E \mid \mathcal{A}(X) = b\}$ is nonempty and consider for some $Z \in E$ the optimization problem*

$$\begin{aligned} \Pi^{\mathcal{W}}_{\mathcal{L}_b}(Z) := & \underset{X\in E}{argmin} && \|\mathcal{W}[X - Z]\| \\ & subject\ to && \mathcal{A}(X) = b \end{aligned} \tag{7}$$

then the solution is given by

$$\Pi^{\mathcal{W}}_{\mathcal{L}_b}(Z) = Z - \mathcal{W}^{-2}\mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b]. \tag{8}$$

Proof. Note that in comparison to Lemma 2.1 the norm in our example above can be regarded as the result of a weighted inner product. In consequence the characterization from Lemma 2.1 for our weighted projections changes: If $X \in \mathcal{L}_b$ then $X = \Pi^{\mathcal{W}}_{\mathcal{L}_b}(Z)$ if and only if $\langle \mathcal{W}[Z-X], \mathcal{W}[Y-X]\rangle \le 0 \ \forall Y \in \mathcal{L}_b$. Note that $(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+\mathcal{A} = \mathcal{A}$ holds true,which implies, since $\mathcal{L}_b$ is nonempty, that $X = Z - \mathcal{W}^{-2}\mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b] \in \mathcal{L}_b$ and therefore the following computation for some arbitrary $Y \in \mathcal{L}_b$ shows

$$\begin{aligned} & \langle \mathcal{W}[Z - X], \mathcal{W}[Y - X]\rangle \\ =& \langle \mathcal{W}[\mathcal{W}^{-2}\mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b]], \mathcal{W}[Y - Z + \mathcal{W}^{-2}\mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b]]\rangle \\ =& \langle \mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b], Y - Z + \mathcal{W}^{-2}\mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b]\rangle \\ =& [\mathcal{A}(Z) - b]^T(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+\mathcal{A}[Y - Z + \mathcal{W}^{-2}\mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b]] \\ =& [\mathcal{A}(Z) - b]^T[(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+\mathcal{A}(Y - Z) + (\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b]] \\ =& [\mathcal{A}(Z) - b]^T[(\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[b - \mathcal{A}(Z)] + (\mathcal{A}\mathcal{W}^{-2}\mathcal{A}^*)^+[\mathcal{A}(Z) - b]] \\ =& 0 \end{aligned} \tag{9}$$

which completes the proof. ■

Example 2.3 (Dual weighted Projection onto an affine Subspace)**.** *Let $\mathcal{W} : E \to E$ be some self-adjoint invertible Linear Operator. Here we consider the (nonempty by definition) affine subspace $\mathcal{L}^\perp - C := \{Y \in E \mid \exists y \in \mathbb{R}^m : \ \mathcal{A}^*(y) - Y = C\}$ and for some $Z \in E$ the optimization problem*

$$\begin{aligned} \Pi^{\mathcal{W}}_{\mathcal{L}^\perp - C}(Z) := & \underset{Y\in E}{argmin} && \|\mathcal{W}[Y - Z]\| \\ & \underset{y\in\mathbb{R}^m}{subject\ to} && \mathcal{A}^*(y) - Y = C \end{aligned} \tag{10}$$

which attains the optimal solution

$$\Pi^{\mathcal{W}}_{\mathcal{L}^\perp - C}(Z) = \mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)] - C. \tag{11}$$

Proof. It is obvious that $Y = \mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)] - C$ satisfies $Y \in \mathcal{L}^\perp - C$ Since we have again "changed" the inner product it follows from the previous Lemma, that $Y = \Pi^{\mathcal{W}}_{\mathcal{L}^\perp - C}(Z)$ if and only if $\langle \mathcal{W}[Z-Y], \mathcal{W}[X-Y]\rangle \leq 0\ \forall X \in \mathcal{L}^\perp - C$. Therefore we choose $X = \mathcal{A}(x) - C \in \mathcal{L}^\perp - C$ for some arbitrary $x \in \mathbb{R}^m$ and derive

$$\begin{aligned}
&\langle \mathcal{W}[Z-Y], \mathcal{W}[X-Y]\rangle \\
=&\langle \mathcal{W}[C+Z-\mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)]], \mathcal{W}[X-\mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)]+C]\rangle \\
=&\langle \mathcal{W}^2[C+Z-\mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)]], \mathcal{A}^*(x)-\mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)]\rangle \\
=&[\mathcal{A}\mathcal{W}^2[C+Z-\mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)]]]^T[x-(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)]] \\
=&[\mathcal{A}\mathcal{W}^2[C+Z] - \underbrace{\mathcal{A}\mathcal{W}^2\mathcal{A}^*(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+\mathcal{A}}_{=\mathcal{A}}[\mathcal{W}^2(C+Z)]]^T[x-(\mathcal{A}\mathcal{W}^2\mathcal{A}^*)^+[\mathcal{A}\mathcal{W}^2(C+Z)]] \\
=&0
\end{aligned} \tag{12}$$

which proves our claim. ■

Remark 2.4. *We note that one can derive similar results for the case when $\mathcal{W}$ is not self-adjoint or even when it is singular. In the latter case however we would usually not speak of a projection, but of a weighted least squares (potentially minimal norm) problem.*

Linear subspaces pose the first example of what we are really interested: Convex cones. In fact we will focus mostly on the situation when $\mathcal{C}$ is not only a convex set, but also a cone. The following well established result traces back to a work of Moreau [60] from 1962 and will be fundamental for our further analysis.

Lemma 2.5 (Moreau's Decomposition)**.** *For a nonempty closed convex cone $\mathcal{K} \subset E$ and an arbitrary $Z \in E$, the following two statements are equivalent:*

1. $Z = X + Y$, $X \in \mathcal{K}$, $Y \in \mathcal{K}^P$ *with* $\langle X, Y\rangle = 0$
2. $X = \Pi_{\mathcal{K}}(Z)$ *and* $Y = \Pi_{\mathcal{K}^P}(Z)$

Definition 2.6 (Generalized Absolute Value)**.** *In order to keep our notation short we introduce the generalized absolute value $|.|_{\mathcal{K}} : E \to E$ as*

$$|Z|_{\mathcal{K}} := \Pi_{\mathcal{K}}(Z) - \Pi_{\mathcal{K}^P}(Z) \tag{13}$$

and recover the projections via

$$\Pi_{\mathcal{K}}(Z) = \frac{1}{2}(Z + |Z|_{\mathcal{K}}), \qquad \Pi_{\mathcal{K}^P}(Z) = \frac{1}{2}(Z - |Z|_{\mathcal{K}}) \tag{14}$$

which will be extensively used in the following.

Now obviously the properties from Lemma 2.1 and 2.5 translate from projections to the generalized absolute value, let us collect the ones that we are going to use in the following.

Lemma 2.7 (Properties of the generalized absolute Value)**.** *Let $\mathcal{K} \subset E$ be a nonempty closed convex cone. The generalized absolute value $|.|_{\mathcal{K}} : E \to E$ possesses the following properties:*

- *The norm of the generalized absolute value is equal to the norm of its argument:*

$$\||Z|_{\mathcal{K}}\| = \|Z\| \quad \forall Z \in E$$

- *$|.|_{\mathcal{K}}$ is nonexpansive (Lipschitz continuous with modulo one), and*

$$\begin{aligned} 0 \leq& |\langle |X|_{\mathcal{K}}, Y\rangle - \langle X, |Y|_{\mathcal{K}}\rangle| \leq \ \langle |X|_{\mathcal{K}}, |Y|_{\mathcal{K}}\rangle - \langle X, Y\rangle \\ =& \frac{1}{2}\|X - Y\|^2 - \frac{1}{2}\||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|^2 \quad \forall X, Y \in E \end{aligned} \tag{15}$$

  is satisfied.

- *$|.|_{\mathcal{K}}$ is the derivative of the usually[3] nonconvex function $Z \mapsto \frac{1}{2}\langle Z, |Z|_{\mathcal{K}}\rangle$, in particular*

$$0 \leq |\frac{1}{2}\langle X, |X|_{\mathcal{K}}\rangle + \frac{1}{2}\langle Y, |Y|_{\mathcal{K}}\rangle - \langle |X|_{\mathcal{K}}, Y\rangle| \leq \frac{1}{2}\|X - Y\|^2 \quad \forall X, Y \in E \tag{16}$$

  holds true.

- *$|.|_{\mathcal{K}}$ is positively homogeneous $|\lambda Z|_{\mathcal{K}} = \lambda |Z|_{\mathcal{K}} \;\; \forall Z \in E \;$ and $\; \forall \lambda \geq 0$*

Proof. The norm property follows directly from Moreaus decomposition. Nonexpansiveness follows by rewriting the inequalities $\langle \Pi_{\mathcal{K}}(X), \Pi_{\mathcal{K}^P}(Y)\rangle \leq 0$ and $\langle \Pi_{\mathcal{K}}(Y), \Pi_{\mathcal{K}^P}(X)\rangle \leq 0$ in terms of the generalized absolute value and proper rearrangement. The inequalities for differentiability follow directly from rewriting the inequalities $\|X - \Pi_{\mathcal{K}}(X)\|^2 \leq \|X - \Pi_{\mathcal{K}}(Y)\|^2$ and $\|X - \Pi_{\mathcal{K}^P}(X)\|^2 \leq \|X - \Pi_{\mathcal{K}^P}(Y)\|^2$. ■

Of course there exist many other projection properties, most of which can be found in Zarantonellos early work [98]. For example the following (which we will not explicitly use and therefore not prove, but may still be of interest to other researchers).

Lemma 2.8 (Potentially useful equalities)**.** *Let $\mathcal{K} \subset E$ be a nonempty closed convex cone. The generalized absolute value $|.|_{\mathcal{K}} : E \to E$ satisfies the following properties:*

- $\frac{1}{2}\|X - Y\|^2 + \frac{1}{2}\||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|^2 = \|\Pi_{\mathcal{K}}(X) - \Pi_{\mathcal{K}}(Y)\|^2 + \|\Pi_{\mathcal{K}^P}(X) - \Pi_{\mathcal{K}^P}(Y)\|^2$
- $\|X - Y\|^2 - \||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|^2 = 4\langle \Pi_{\mathcal{K}}(X) - \Pi_{\mathcal{K}}(Y), \Pi_{\mathcal{K}^P}(X) - \Pi_{\mathcal{K}^P}(Y)\rangle$
- $\|X - Y\|^2 - \||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|^2 = \||X|_{\mathcal{K}} + |Y|_{\mathcal{K}}\|^2 - \|X + Y\|^2$

[3]Except for some generic case, for example when $\mathcal{K} = \mathbb{R}$, then $|z|_{\mathbb{R}} z = z^2$, which is obviously convex.

2.2 The Reduced Lagrangian

Now that we have recalled the basics about projections and the generalized absolute value, let us go ahead and start by applying Moreau's Decomposition together with our new notation (14) to the sufficient optimality conditions (4). This yields the conditions

$$\begin{aligned} &\mathcal{A}^*(y) - \tfrac{1}{2}(Z - |Z|_{\mathcal{K}}) = C \\ &\tfrac{1}{2}\mathcal{A}(Z + |Z|_{\mathcal{K}}) = b \\ &Z \in E,\ y \in \mathbb{R}^m \end{aligned} \tag{17}$$

which we are going to reformulate further: By rewriting the first equation of (17) as $\frac{1}{2}|Z|_{\mathcal{K}} = C - \mathcal{A}^*(y) + \frac{1}{2}Z$ and substituting this into the second equation yields $\mathcal{A}[Z + C - \mathcal{A}^*(y)] = b$. If we assume that $\mathcal{A}$ is a surjective operator, which we do without loss of generality for the rest of this thesis, we can express this equation in terms of $y \in \mathbb{R}^m$ as $y = (\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(Z+C) - b]$. Eliminating y from the first equation of (17) gives the condition

$$\nabla f(Z) := \quad \tfrac{1}{2}(Z - |Z|_{\mathcal{K}}) + C - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(Z + C) - b] \stackrel{!}{=} 0 \tag{18}$$

The fact that the right hand side in the definition (18) is the gradient of some function f will be established in Lemma 2.10 below. The next result concerns the relation with (4):

Corollary 2.9. *The connection between the KKT-conditions* (4) *and the function* $\nabla f : E \to E$ *is the following:*

- *If* $X^{(*)}, Y^{(*)} \in E,\ y^{(*)} \in \mathbb{R}^m$ *satisfy the KKT-conditions* (4) *then* $\nabla f(X^{(*)} + Y^{(*)}) = 0$
- *If* $Z^{(*)} \in E$ *satisfies* $\nabla f(Z^{(*)}) = 0$ *then* $X^{(*)} := \Pi_{\mathcal{K}}(Z^{(*)}),\ Y^{(*)} := \Pi_{\mathcal{K}^P}(Z^{(*)})$ *and* $y^{(*)} := (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}[C + \Pi_{\mathcal{K}^P}(Z^{(*)})]$ *satisfy the conic KKT-conditions* (4)

Proof. The first part follows directly from the derivation and definition of ∇f. To make the second part clear, we note that for all $Z \in E$ we can rewrite the definition of ∇f in (18) as

$$\begin{aligned} \nabla f(Z) &= \underbrace{\tfrac{1}{2}(Z - |Z|_{\mathcal{K}})}_{=\Pi_{\mathcal{K}^P}(Z)} + C - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(\underbrace{Z}_{=\Pi_{\mathcal{K}}(Z) + \Pi_{\mathcal{K}^P}(Z)} + C) - b] \\ &= (\mathcal{I} - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})[\Pi_{\mathcal{K}^P}(Z)) + C] - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(\Pi_{\mathcal{K}}(Z)) - b] \end{aligned} \tag{19}$$

and therefore obtain that by orthogonality

$$\|\nabla f(Z)\|^2 = \left\|\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(\Pi_{\mathcal{K}}(Z)) - b]\right\|^2 + \left\|(\mathcal{I} - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})[\Pi_{\mathcal{K}^P}(Z)) + C]\right\|^2 \tag{20}$$

holds true. Any $Z^{(*)} \in E$ satisfying $\nabla f(Z^{(*)}) = 0$ then obviously implies $\mathcal{A}(\Pi_{\mathcal{K}}(Z^{(*)})) = b$ and $\mathcal{A}^*(y^{(*)}) - \Pi_{\mathcal{K}^P}(Z^{(*)}) = C$ which proves our claim. ■

In order to investigate properties of ∇f we shorten our notation by defining the generalized Householder transformation

$$\mathcal{H} := \quad \mathcal{I} - 2\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A} \tag{21}$$

where $\mathcal{I}$ denotes the identity operator. By using the above definition as well as the following definition for the constant term

$$\mathcal{R} := \quad C - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(C) - b]$$

we see that

$$\nabla f(Z) = \quad \frac{1}{2}\mathcal{H}[Z] - \frac{1}{2}|Z|_{\mathcal{K}} + \mathcal{R} \tag{22}$$

holds true. Note that $\mathcal{H}$ is, with respect to the inner product $\langle .,. \rangle$, self-adjoint and orthogonal (and therefore really is a generalization of Householder matrices with $\mathcal{H}^2 = \mathcal{I}$) and reflects/mirrors the term

$$\mathcal{R} = (\mathcal{I} - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})[C] + \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}b$$

onto

$$\mathcal{H}[\mathcal{R}] = (\mathcal{I} - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})[C] - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}b \tag{23}$$

which we will use shortly. Keep in mind that in any practical implementation we would usually not form any of the above terms explicitly. In fact we can even go a step further by avoiding to form the operators $\mathcal{A}$, $\mathcal{A}^*$ and $(\mathcal{A}\mathcal{A}^*)^{-1}$. Although these operators are of course linear and therefore matrix-representable, it might be (and often is) computationally beneficial to just implement their application $\mathcal{A}(X)$, $\mathcal{A}^*(y)$ and $(\mathcal{A}\mathcal{A}^*)^{-1}y$ to arbitrary points $X \in E,\ y \in \mathbb{R}^m$. While this makes mathematically no difference, the computational effort may decrease dramatically if the operators are of special form, for example due to some sort of Kronecker-, low-rank-, sparsity- and/or symmetry-structure. Another observation regarding the evaluation complexity that we can and should make is the following: The evaluation of ∇f or of the "reflected" function $\mathcal{H}\nabla f$ only involves one and not two applications of $\mathcal{A}$, $\mathcal{A}^*$ and $(\mathcal{A}\mathcal{A}^*)^{-1}$ each, if done efficiently. While for ∇f this is easily seen from equation (18), let us derive the formula for $\mathcal{H}\nabla f$ here explicitly. From using (22), orthogonality of $\mathcal{H}$ and (23) it follows that

$$\begin{aligned}\mathcal{H}[\nabla f(Z)] =& \mathcal{H}[\tfrac{1}{2}\mathcal{H}[Z] - \tfrac{1}{2}|Z|_{\mathcal{K}} + \mathcal{R}] = \tfrac{1}{2}Z - \mathcal{H}[\tfrac{1}{2}|Z|_{\mathcal{K}}] + \mathcal{H}[\mathcal{R}] \\ =& \tfrac{1}{2}(Z - |Z|_{\mathcal{K}}) + C + \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(|Z|_{\mathcal{K}} - C) - b]\end{aligned} \tag{24}$$

holds true for arbitrary $Z \in E$. Above formula becomes especially important in our algorithmic section, where the operator $\mathcal{H}\nabla f$ will be one of our main workhorses, making an efficient evaluation essential. Now before we investigate further properties of ∇f, we justify our notation. The following lemma shows that ∇f is in fact a gradient and additionally supplies us with a connection to the optimal value of our initial problems (1) respectively (2).

Lemma 2.10. *$\nabla f(Z)$ is the gradient of the differentiable, but usually nonconvex function $f : E \to \mathbb{R}$*

$$f(Z) := \tfrac{1}{4}\langle Z, \mathcal{H}Z - |Z|_{\mathcal{K}}\rangle + \langle \mathcal{R}, Z\rangle + const \tag{25}$$

which we will call reduced Lagrangian, explaining our notation in (18). *For the choice $const = \frac{1}{2}\|C\|^2 - \frac{1}{2}[\mathcal{A}(C)-b]^T(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(C)-b]$ and any $Z^{(*)} \in E$ satisfying $\nabla f(Z^{(*)}) = 0$ we recover the optimal value of* (1) *via*

$$f(Z^{(*)}) = \langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle \tag{26}$$

Proof. The form (25) follows immediately from equation (22) in combination with the antiderivative of the generalized absolute value (in Lemma 2.7) and linearity of differentiation. Note that an $\epsilon-\delta$-type of proof can also be easily derived from Lemma 2.14 below. We can rewrite (25) by using (22) and obtain the identity

$$f(Z) = \tfrac{1}{2}\langle Z, \nabla f(Z)\rangle + \tfrac{1}{2}\langle \mathcal{R}, Z\rangle + const \quad \forall Z \in E. \tag{27}$$

For $Z^{(*)} \in E$ satisfying $\nabla f(Z^{(*)}) = 0$ we therefore obtain, by using first the definition of $\mathcal{R}$ and then corollary 2.9, that the equalities

$$\begin{aligned} \langle \mathcal{R}, \Pi_{\mathcal{K}}(Z^{(*)})\rangle &= \langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle - \langle \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(C)-b], \Pi_{\mathcal{K}}(Z^{(*)})\rangle \\ &= \langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle - [\mathcal{A}(C)-b]^T(\mathcal{A}\mathcal{A}^*)^{-1}b \end{aligned} \tag{28}$$

and, for $y^{(*)} = (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}[C + \Pi_{\mathcal{K}^P}(Z^{(*)})]$ from corollary 2.9.,

$$\begin{aligned} \langle \mathcal{R}, \Pi_{\mathcal{K}^P}(Z^{(*)})\rangle &= \langle C - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(C)-b], \mathcal{A}^*(y^{(*)}) - C\rangle \\ &= b^T y^{(*)} - \langle C - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(C)-b], C\rangle \end{aligned} \tag{29}$$

hold true. This implies together with (27)

$$\begin{aligned} f(Z^{(*)}) &= \tfrac{1}{2}\langle Z^{(*)}, \underbrace{\nabla f(Z^{(*)})}_{=0}\rangle + \tfrac{1}{2}\langle \mathcal{R}, \underbrace{Z^{(*)}}_{=\Pi_{\mathcal{K}}(Z^{(*)})+\Pi_{\mathcal{K}^P}(Z^{(*)})}\rangle + const \\ &= \tfrac{1}{2}\langle \mathcal{R}, \Pi_{\mathcal{K}}(Z^{(*)}) + \Pi_{\mathcal{K}^P}(Z^{(*)})\rangle + const \overset{(28),(29)}{=} \tfrac{1}{2}\langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle + \underbrace{\tfrac{1}{2}b^T y^{(*)}}_{=\frac{1}{2}\langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle} \end{aligned} \tag{30}$$

where we used that the choice of *const* can be rewritten as

$$\begin{aligned} const &= \frac{1}{2}\|C\|^2 - \frac{1}{2}[\mathcal{A}(C)-b]^T(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(C)-b] \\ &= [\mathcal{A}(C)-b]^T(\mathcal{A}\mathcal{A}^*)^{-1}b + \langle C - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(C)-b], C\rangle \end{aligned}$$

which proves our claim. ■

As we have seen, solving (1) respectively (2) is equivalent to finding a zero of ∇f. Therefore it is obvious that it is also equivalent to finding a zero of $\mathcal{H}\nabla f$. Now this trivial observation actually makes developing algorithms with global convergence behavior much simpler: Since f is non-convex (except for some generic problems (1)) any zero of ∇f will (usually) neither be a local minimum nor a local maximum, but a saddle point of f. Therefore the gradient does not give us much information about the the direction towards this(these) saddle point(s). Quite remarkable however, the reflected gradient $\mathcal{H}\nabla f$ will give us a sensible direction. This is due to the fact that the operator $\mathcal{H}\nabla f$ is not only Lipschitz continuous but in fact firmly nonexpansive (or 1-cocoercive). Recall that an operator $g : E \to E$ is called firmly nonexpansive if the equation

$$\|g(X) - g(Y)\|^2 \leq \langle g(X) - g(Y), X - Y\rangle$$

holds true for all $X, Y \in E$. Above claims and some other properties of ∇f are summarized and proved next.

Proposition 2.11 (Properties of ∇f)**.** *Let $\nabla f : E \to E$ be defined as in* (18) *respectively* (22)*. The following holds:*

- *∇f is just an orthogonal factor ($\mathcal{H}$) away from being firmly nonexpansive*

$$\|\nabla f(X) - \nabla f(Y)\|^2 \leq \langle \nabla f(X) - \nabla f(Y), \mathcal{H}[X - Y]\rangle \quad \forall X, Y \in E \tag{31}$$

  and together with the orthogonality and self-adjointness of $\mathcal{H}$ this implies that the reflected gradient $\mathcal{H}\nabla f$ is firmly nonexpansive.

- *∇f is nonexpansive (Lipschitz continuous with modulo one), i. e.*

$$\|\nabla f(X) - \nabla f(Y)\| \leq \|X - Y\| \quad \forall X, Y \in E \tag{32}$$

  holds true.

- *On the other hand we also have*

$$-\langle \nabla f(X) - \nabla f(Y), |X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\rangle \leq \|\nabla f(X) - \nabla f(Y)\|^2 \quad \forall X, Y \in E \tag{33}$$

- *And finally,*

$$0 \leq \frac{1}{2}\|X - Y\| - \frac{1}{2}\||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\| \leq \|\nabla f(X) - \nabla f(Y)\| \quad \forall X, Y \in E. \tag{34}$$

Proof. By using (22), we can write $\nabla f(X) - \nabla f(Y) = \frac{1}{2}\mathcal{H}[X - Y] - \frac{1}{2}(|X|_{\mathcal{K}} - |Y|_{\mathcal{K}})$. The first and third inequality then follow from

$$\begin{aligned}
&\|\nabla f(X) - \nabla f(Y)\|^2 - \langle \nabla f(X) - \nabla f(Y), \mathcal{H}[X - Y]\rangle \\
=&\langle \nabla f(X) - \nabla f(Y), \nabla f(X) - \nabla f(Y) - \mathcal{H}[X - Y]\rangle \\
=&\langle \tfrac{1}{2}\mathcal{H}[X - Y] - \tfrac{1}{2}(|X|_{\mathcal{K}} - |Y|_{\mathcal{K}}), -\tfrac{1}{2}\mathcal{H}[X - Y] - \tfrac{1}{2}(|X|_{\mathcal{K}} - |Y|_{\mathcal{K}})\rangle \\
=&\underbrace{-\frac{1}{4}\|X - Y\|^2 + \frac{1}{4}\||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|^2}_{\leq 0}
\end{aligned} \tag{35}$$

and

$$\begin{aligned}
&\|\nabla f(X) - \nabla f(Y)\|^2 + \langle \nabla f(X) - \nabla f(Y), |X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\rangle \\
=&\langle \nabla f(X) - \nabla f(Y), \nabla f(X) - \nabla f(Y) + (|X|_{\mathcal{K}} - |Y|_{\mathcal{K}})\rangle \\
=&\langle \tfrac{1}{2}\mathcal{H}[X - Y] - \tfrac{1}{2}(|X|_{\mathcal{K}} - |Y|_{\mathcal{K}}), \tfrac{1}{2}\mathcal{H}[X - Y] + \tfrac{1}{2}(|X|_{\mathcal{K}} - |Y|_{\mathcal{K}})\rangle \\
=&\underbrace{\frac{1}{4}\|X - Y\|^2 - \frac{1}{4}\||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|^2}_{\geq 0}.
\end{aligned} \tag{36}$$

The second inequality is an obvious implication of the well known Cauchy-Schwarz inequality (i.e. $\langle X, Y\rangle \leq \|X\|\,\|Y\| \forall X, Y \in E$). The last inequality is implied by the inverse triangle inequality (i.e. $|\,\|X\| - \|Y\|\,| \leq \|X - Y\| \quad \forall X, Y \in E$) and Lipschitz continuity of $|.|_{\mathcal{K}}$. ∎

Let us emphasize the importance above proposition and especially inequality (31). One important implication concerns the shape of the set of points that we are looking for, i.e. the set of all zeros of ∇f. While it is clear that the set of primal or dual optimal solutions of (1) respectively (2) are convex sets, this is not strictly obvious for the set of zeros of ∇f. Note that for example sine and cosine are globally 1-Lipschitz, but their sets of zeros are not convex. Therefore (32) is not sufficient, but by using (31) we can show the following result.

Lemma 2.12. *The inverse image of zero*

$$(\nabla f)^{-1}(0) := \{Z \in E \mid \nabla f(Z) = 0\} \tag{37}$$

is a closed and convex set in E.

Proof. Since ∇f is continuous, the inverse image of the closed set $\{0\} \subset E$ must be closed. Let now $X^{(*)}, Y^{(*)} \in E$ satisfy $\nabla f(X^{(*)}) = \nabla f(Y^{(*)}) = 0$. For $Z^{(*)} := \frac{1}{2}(X^{(*)} + Y^{(*)})$ the computation

$$\begin{aligned} 2\left\|\nabla f(Z^{(*)})\right\|^2 &= \left\|\nabla f(Z^{(*)}) - \nabla f(X^{(*)})\right\|^2 + \left\|\nabla f(Z^{(*)}) - \nabla f(Y^{(*)})\right\|^2 \\ &\leq \langle \nabla f(Z^{(*)}) - \nabla f(X^{(*)}), \mathcal{H}[Z^{(*)} - X^{(*)}]\rangle + \langle \nabla f(Z^{(*)}) - \nabla f(Y^{(*)}), \mathcal{H}[Z^{(*)} - Y^{(*)}]\rangle \\ &= \langle \nabla f(Z^{(*)}), \mathcal{H}[\tfrac{1}{2}(X^{(*)} + Y^{(*)}) - X^{(*)}]\rangle + \langle \nabla f(Z^{(*)}), \mathcal{H}[\tfrac{1}{2}(X^{(*)} + Y^{(*)}) - Y^{(*)}]\rangle \\ &= 0 \end{aligned} \tag{38}$$

where we used (31) in the second line, shows that $Z^{(*)} \in E$ also satisfies $\nabla f(Z^{(*)}) = 0$. Due to continuity of ∇f this completes the proof. ■

Note that above lemma implies a well defined projection operator onto the set $(\nabla f)^{-1}(0)$ whenever it is nonempty, which is a fact that we can exploit for asymptotic convergence behavior later on. There is one ambiguity that we address in the remaining subsection: Most of our convergence results only rely on the firm-nonexpansiveness of $\mathcal{H}\nabla f$. Now why did we not simply define an operator $g := \mathcal{H}\nabla f$, look for a zero $Z^{(*)} \in E$ of g and never talk about f again? After all, the objective values can be recovered not via a "weirdly" defined function f but simply by considering $\langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle$ for example. We could have saved so many pages. Well... no! First of all: in practice we never really find any zero of g. What we do find is a point $Z \in E$ with $g(Z) \approx 0$. Unfortunately the symbol "$\approx$" does not hold any implications about the magnitude of $\|Z - Z^{(*)}\|$, which could in fact be arbitrarily large. What does this imply for the objective value $\langle C, \Pi_{\mathcal{K}}(Z)\rangle$? Are we close to the optimal value? The painful answer is: we don't know. Meanwhile $f(Z)$ is a much more educated guess for the optimal value as we are going to show next. Our analysis is of course based on essentially a Taylor expansion. First of all note that, by the definition (25) and equality (22), the following equality

$$f(X) - f(Y) - \langle \nabla f(Y), X - Y\rangle = \tfrac{1}{2}\langle \nabla f(X) - \nabla f(Y), X - Y\rangle + \tfrac{1}{4}(\langle |Y|_{\mathcal{K}}, X\rangle - \langle |X|_{\mathcal{K}}, Y\rangle) \tag{39}$$

holds true for all $X, Y \in E$. Moreover we see that the following holds true (which requires some very rough inequalities):

Lemma 2.13 (Sensitivity of f)**.** *Similar to the quadratic case we have*

$$|f(X) - f(Y)| \leq \tfrac{1}{2} \|X - Y\| \left(\|\nabla f(X) + \nabla f(Y)\| + \|\nabla f(X) - \nabla f(Y)\|\right) \tag{40}$$

for all $X, Y \in E$. which implies that for any $Z^{(*)} \in E$ with $\nabla f(Z^{(*)}) = 0$ and any $Z \in E$ the inequality

$$|f(Z) - f(Z^{(*)})| \leq \|\nabla f(Z)\| \left\|Z - Z^{(*)}\right\| \tag{41}$$

holds true.

Note that both factors of the right hand side of (41) tend to zero as Z approaches $Z^{(*)}$. When the constant part of f is defined according to Lemma 2.10, then $f(Z^{(*)}) = \langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle$ and the bound (41) provides a much better estimate for the optimal objective value than the bound $O(\|Z - Z^{(*)}\|)$ of a solution Z somewhere near the optimal solution $Z^{(*)}$. In fact, the definition of the function f by itself and the 1-Lipschitz continuity of its gradient already allow the bound $\|Z - Z^{(*)}\|^2$ based on the Taylor expansion. This improved bound is possible because of the fact that both primal and dual variables are condensed in Z. The bound (41) typically improves the quadratic bound above since often $\|\nabla f(Z)\| \ll \|Z - Z^{(*)}\|$ so that the bound (41) generally is also tighter.

Proof. Note that the second part follows instantly for $X = Z$ and $Y = Z^{(*)}$. In order to prove the first part we can rewrite (39) as

$$f(X) - f(Y) = \frac{1}{2}\langle \nabla f(X) + \nabla f(Y), X - Y\rangle + \frac{1}{4}(\langle |Y|_{\mathcal{K}}, X\rangle - \langle |X|_{\mathcal{K}}, Y\rangle) \tag{42}$$

and take the absolute value on both sides. The triangle inequality implies

$$|f(X) - f(Y)| \leq \tfrac{1}{2}|\langle \nabla f(X) + \nabla f(Y), X - Y\rangle| + \tfrac{1}{4}|\langle |Y|_{\mathcal{K}}, X\rangle - \langle |X|_{\mathcal{K}}, Y\rangle|$$

where the first summand can be upper bounded via the Cauchy-Schwarz inequality, i.e.

$$\tfrac{1}{2}|\langle \nabla f(X) + \nabla f(Y), X - Y\rangle| \leq \tfrac{1}{2} \|\nabla f(X) + \nabla f(Y)\| \|X - Y\| \tag{43}$$

and the second summand by using inequalities (15), (34) and again (15)

$$\begin{aligned}
&\tfrac{1}{4}|\langle |Y|_{\mathcal{K}}, X\rangle - \langle |X|_{\mathcal{K}}, Y\rangle| \\
\leq&\tfrac{1}{8} \|X - Y\|^2 - \||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|^2 \\
=&\tfrac{1}{4}(\|X - Y\| + \||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|)(\tfrac{1}{2} \|X - Y\| - \tfrac{1}{2} \||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|) \\
\leq&\tfrac{1}{4}(\|X - Y\| + \||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\|) \|\nabla f(X) - \nabla f(Y)\| \\
\leq&\tfrac{1}{2} \|X - Y\| \|\nabla f(X) - \nabla f(Y)\|
\end{aligned} \tag{44}$$

which implies the first part and therefore our proof. ■

To conclude our earlier argument, note that, if the constant part of f has been chosen according to Lemma 2.10, inequality (41) implies that we can approximate the optimal value $f(Z^{(*)}) = \langle C, \Pi_{\mathcal{K}}(Z^{(*)})\rangle$ of (1) quite well with $f(Z)$ whenever its gradients norm is sufficiently small. However we also called the used inequalities "quite rough" and in fact we will use the following finer inequalities in our analysis later on.

Lemma 2.14. *Let $f : E \to \mathbb{R}$ be defined as in Lemma 2.10. then*

$$\begin{aligned} &\tfrac{1}{2}\langle(\mathcal{I}-\mathcal{H})(\nabla f(X)-\nabla f(Y)), X-Y\rangle + \tfrac{1}{2}\left\|\nabla f(X)-\nabla f(Y)\right\|^2 \\ \leq & \qquad f(X)-f(Y)-\langle\nabla f(Y), X-Y\rangle \\ \leq & \tfrac{1}{2}\langle(\mathcal{I}+\mathcal{H})(\nabla f(X)-\nabla f(Y)), X-Y\rangle - \tfrac{1}{2}\left\|\nabla f(X)-\nabla f(Y)\right\|^2 \end{aligned} \tag{45}$$

hold true for all $X, Y \in E$.

Proof. This is an immediate consequence of equality (39) above and Lemma 2.7.: We use $\langle|X|_{\mathcal{K}}, Y\rangle - \langle X, |Y|_{\mathcal{K}}\rangle \geq \langle X, Y\rangle - \langle|X|_{\mathcal{K}}, |Y|_{\mathcal{K}}\rangle$ for the first, $\langle|X|_{\mathcal{K}}, Y\rangle - \langle X, |Y|_{\mathcal{K}}\rangle \leq \langle|X|_{\mathcal{K}}, |Y|_{\mathcal{K}}\rangle - \langle X, Y\rangle$ for the second inequality (both follow from (15)) and the equality

$$\begin{aligned} \tfrac{1}{4}(\langle|X|_{\mathcal{K}}, |Y|_{\mathcal{K}}\rangle - \langle X, Y\rangle) &= \tfrac{1}{8}\left\|X-Y\right\|^2 - \tfrac{1}{8}\left\||X|_{\mathcal{K}} - |Y|_{\mathcal{K}}\right\|^2 \\ &= \tfrac{1}{2}\langle\nabla f(X)-\nabla f(Y), \mathcal{H}[X-Y]\rangle - \tfrac{1}{2}\left\|\nabla f(X)-\nabla f(Y)\right\|^2 \end{aligned}$$

which we have essentially already seen in the proof of proposition 2.11. ■

When it comes to a second order Taylor expansion, we can derive a partial quadratic expansion from (39) via

$$\begin{aligned} &f(X)-f(Y)-\langle\nabla f(Y), X-Y\rangle - \tfrac{1}{4}\langle X-Y, \mathcal{H}(X-Y)\rangle \\ =& -\tfrac{1}{4}\langle|X|_{\mathcal{K}}, X\rangle - \tfrac{1}{4}\langle|Y|_{\mathcal{K}}, Y\rangle + \tfrac{1}{2}\langle|Y|_{\mathcal{K}}, X\rangle \end{aligned} \tag{46}$$

where the absolute value of the right-hand-side is bounded above by $\frac{1}{4}\left\|X-Y\right\|^2$ according to (16). The reason why it is difficult to obtain better quadratic models lies in our generalized absolute value $|.|_{\mathcal{K}}$, which is Lipschitz-continuous, but in general not differentiable everywhere. In order to overcome this we need a concept of generalized derivatives, specifically we are going to use Clarke's approach [11] here and remind the reader of this concept in the next subsection.

2.3 Generalized Derivatives

This section recalls the concept of generalized derivatives in the sense of Clarke [11]. It is not meant as a general discussion, nor is it in any sense complete: We will keep everything as simple as possible, while maintaining suitability for our application. Ideally we would simply define Clarke's generalized Jacobian, but unfortunately there is a technical issue: We do work in a finite dimensional Euclidean setting, i.e. a setting that is of course equivalent, but not strictly the same as the real coordinate space $\mathbb{R}^n$, which is the scenario that Clarke considered originally. Let us sketch the differences by pointing out that in the real coordinate space, we commonly understand the points as vectors and the Jacobian or the Hessian as matrices. For us however even the points can for example be matrices, implying that function derivatives at such points are not truly matrices but linear operators. These operators are of course matrix representable, but we like to avoid the explicit isomorphism as it is considered superfluous and less elegant. Below, we therefore summarize the main results of Clarke in the abstract setting of a Euclidean space E suited for the framework considered in this thesis. Before generalizing the concept of derivatives, it is probably best to start with a definition fixing the common concept of derivatives for this subsection. Here we will restrict ourselves to functions where domain and co-domain coincide, as this will be sufficient for our needs.

Definition 2.15. *A function $F : E \to E$ is called Fréchet- (or F-) differentiable at a point $Z \in E$ if there exists a bounded linear Operator $\nabla F(Z) : E \to E$ such that*

$$\lim_{\Delta Z \to 0} \frac{\|F(Z + \Delta Z) - F(Z) - \nabla F(Z)[\Delta Z]\|}{\|\Delta Z\|} = 0 \tag{47}$$

We will call $\nabla F(Z)$ the Fréchet-derivative. We will say that F is continuously differentiable in $Z \in E$ if there exists an open neighborhood $U \subset E$ of Z on which F is F-differentiable and the Fréchet-derivative is continuous, i.e. the limit $\lim_{\Delta Z \to 0} \nabla f(Z + \Delta Z)$ exists and coincides with $\nabla f(Z)$.

We kept the term "bounded" in this definition to avoid confusion, but it is superfluous, because E is finite dimensional. Also keep in mind that for Lipschitz-continuous functions in a finite dimensional Euclidean space, the concepts of total- or Gateaux- differentiability are essentially equivalent to F-differentiability. For our further analysis, Lipschitz-continuous functions will play a crucial role due to the famous theorem of Rademacher [73]:

Theorem 2.16 (Rademacher)**.** *Let $F : E \to E$ be a (locally) Lipschitz-continuous function. Then F is Fréchet-differentiable almost everywhere, i.e. the set $\Omega_F \subset E$ where F is not Fréchet-differentiable has Lebesgue measure zero.*

Now in order to generalize derivatives, we start by defining the set

$$\mathcal{L}(E, E) := \{\mathcal{V} : E \to E \mid \mathcal{V} \text{ is a bounded linear operator}\} \tag{48}$$

of bounded linear operators. Note that the F-derivative can be seen as a well defined function $\nabla F : E \setminus \Omega_F \to \mathcal{L}(E, E)$. The idea of generalized derivatives is to now "smooth" the gaps Ω_F in a meaningful fashion. One way to accomplish this, is by defining a set-valued function $\partial F : E \to \mathbb{P}(\mathcal{L}(E, E))$, where $\mathbb{P}(\mathcal{L}(E, E))$ denotes the power set of $\mathcal{L}(E, E)$ in the following way:

Definition 2.17 (Generalized Derivatives)**.** *Let $F : E \to E$ be a (locally) Lipschitz-continuous function. For $Z \in E$ the set*

$$\partial_B F(Z) := \{G \in \mathcal{L}(E, E) \mid \exists \{Z^{(k)}\}_{k \in \mathbb{N}} \subset E \setminus \Omega_F \text{ with } Z^{(k)} \to Z,\ \nabla F(Z^{(k)}) \to G\} \tag{49}$$

is called the Bouligand- (or B-) derivative of F and

$$\partial F(Z) := conv(\partial_B F(Z)) \tag{50}$$

Clarke's generalized derivative, where conv denotes the convex hull.

Remark 2.18. *Note that our definition is essentially a special case* [4] *of Clarke's generalized Jacobian (c.f. [11] definition 2.6.1): Since there exists an isometric isomorphism* $vec : E \to \mathbb{R}^{\dim(E)}$*, i.e. vec is an isomorphism that satisfies* $\langle X, Y\rangle = vec(X)^T vec(Y)\ \forall X, Y \in E$*, we can define* $F_{\mathbb{R}} : \mathbb{R}^{\dim(E)} \to \mathbb{R}^{\dim(E)}$ *via*

$$F_{\mathbb{R}}(z) := vec(F(vec^{-1}(z))) \tag{51}$$

and with a slight abuse of notation we obtain

$$\partial F(Z) = vec^{-1} \circ \partial F_{\mathbb{R}}(vec(Z)) \circ vec \tag{52}$$

where for $\partial F_{\mathbb{R}}$ *our definition coincides with the definition of Clarke's generalized Jacobian (i.e.* $\partial F_{\mathbb{R}}$ *can be seen as a set of real quadratic matrices).*

There are a couple of properties that we will repeatedly use. First of all note that for each $Z \in E$ the set $\partial F(Z)$ is non-empty due to Rademacher's theorem, it is convex by definition and it is closed in $\mathcal{L}(E, E)$ because $\partial_B F(Z)$ is closed in $\mathcal{L}(E, E)$ (which follows from its definition). If we define the usual operator norm

$$||\mathcal{V}|| := \sup_{X\in E}\{\|\mathcal{V}[X]\| \mid \|X\| = 1\} \tag{53}$$

then the norm of each element of $\partial F(Z)$ will be upper bounded by the (local) Lipschitz constant, because the F-derivative is locally bounded on its domain. Now we already said that Clarke's generalized derivative "smoothes" the gaps Ω_F where no F-derivative exists. In fact it even does slightly more. For example differentiability of the function F at some point $Z \in E$ is not sufficient for single valuedness of $\partial_B F$ or ∂F as the following example illustrates.

Example 2.19 (see [9])**.** *Define the function* $F : \mathbb{R} \to \mathbb{R}$

$$F(z) = \begin{cases} z^2 \sin(\frac{1}{z}) & \text{if } z \neq 0 \\ 0 & \text{if } z = 0 \end{cases}.$$

which is differentiable everywhere. Especially the derivative in zero is given by $F'(0) = 0$ *but Clarke's generalized derivative is an interval* $\partial F(0) = [-1, 1]$.

In other words, to fully recover the usual F-derivative from the generalized derivative, we need slightly more structure. By definition continuous differentiability is for example sufficient:

[4] We could work with a more general definition of generalized derivatives, namely one where domain and co-domain are not always equal but can differ from each other. However this usually leads to a certain amount of confusion regarding componentwise derivatives. Since our definition is sufficient for all purposes of this thesis, let us just note that Clarke's generalized derivative can also be (and mostly is) defined componentwise for operators $F(Z) := (F_1(Z), ..., F_m(Z))$, such that the inclusion $\partial F(Z) \subset \partial F_1(Z) \times ... \times \partial F_m(Z)$ holds true. For a two dimensional real example $F : \mathbb{R}^2 \to \mathbb{R}^2, F(Z) = (F_1(Z), F_2(Z))$ $F_1 : \mathbb{R}^2 \to \mathbb{R}, F_2 : \mathbb{R}^2 \to \mathbb{R}$, showing that the opposite inclusion is not true in general $\partial F(Z) \neq \partial F_1(Z) \times \partial F_2(Z)$, we refer to example 6.4 in [82]

Lemma 2.20. *If a function $F : E \to E$ is continuously differentiable in an open neighborhood of $Z \in E$ then*

$$\partial_B F(Z) = \partial F(Z) = \{\nabla F(Z)\}. \tag{54}$$

Maybe most important for the design of quasi Newton methods is the following mean value theorem which can be easily seen to hold true from [11] proposition 2.65 (the notation is actually exactly the same, but the interpretation is slightly different):

Proposition 2.21. *Let $F : E \to E$ be Lipschitz continuous and let $X, Y \in E$ be two arbitrary points. Then*

$$F(Y) - F(X) \in conv(\partial F([X, Y]))(Y - X) \tag{55}$$

holds true where $\partial F([X,Y]) := \bigcup_{t\in[0,1]} \partial F((1-t)X + tY)$ denotes the union of all generalized derivatives on the line $[X, Y] := \{(1-t)X + tY \mid t \in [0, 1]\}$.

Let us come back to analyzing the reduced Lagrangian

$$f(Z) = \tfrac{1}{4}\langle Z, \mathcal{H}[Z]\rangle - \tfrac{1}{4}\langle Z, |Z|_{\mathcal{K}}\rangle + \langle Z, \mathcal{R}\rangle + const \tag{56}$$

from before. As we have seen in the previous section f is differentiable with nonexpansive gradient. Due to Rademacher's theorem ∇f is F-differentiable almost everywhere, i.e. the set

$$\Omega_{\nabla f} := \{Z \in E \mid \nabla f \text{ is not F-differentiable in } Z\} \tag{57}$$

of points where ∇f is not F-differentiable has Lebesgue measure zero. Let us define the second derivative of f as the first F-derivative of ∇f whenever it exists i.e.

$$\nabla^2 f(Z) := \nabla(\nabla f(Z)) \tag{58}$$

for all $Z \in E \setminus \Omega_{\nabla f}$. We will say that f is twice (continuously) differentiable in $Z \in E$ whenever ∇f is (continuously) F-differentiable in $Z \in E$. Let us also fix our notation for generalized second derivatives of f for which we define the sets

$$\partial_B^2 f(Z) := \partial_B(\nabla f(Z)) \subset \mathcal{L}(E, E) \tag{59}$$

and

$$\partial^2 f(Z) := \partial(\nabla f(Z)) \subset \mathcal{L}(E, E) \tag{60}$$

for all $Z \in E$. Note that we have the identity $\partial^2 f(Z) = conv(\partial_B^2 f(Z))$ by definition. Now by using the remarkable result 3.3.4 from [67] , we know that the second derivative of f i.e. the F-derivative of ∇f is self-adjoint whenever it exists. We write $(\nabla^2 f(Z))^* = \nabla^2 f(Z)$ meaning that for every $Z \in E \setminus \Omega_{\nabla f}$ the identity

$$\langle \nabla^2 f(Z)[H_1], H_2\rangle = \langle H_1, \nabla^2 f(Z)[H_2]\rangle \tag{61}$$

holds for all $H_1, H_2 \in E$. By definition we immediately conclude the following result.

Proposition 2.22. *Let $Z \in E$ be arbitrary and f as in (56). Then every $\mathcal{M}_B \in \partial_B^2 f(Z)$ and every $\mathcal{M} \in \partial^2 f(Z)$ is self-adjoint. Recall that this implies a real spectrum and the existence of a basis of eigenvectors.*

Above proposition is of special importance for the design of second order methods: Self-adjoint (symmetric) linear systems are generally cheaper (than non-self-adjoint ones) to form, store or solve. Note that while the dimension of $\mathcal{L}(E, E)$ is equal to $\dim(E)^2$, the dimension of the subspace of self-adjoint bounded linear operators is only $\frac{\dim(E)^2+\dim(E)}{2}$. When it comes to solving these systems, they are especially suitable for the application of symmetric iterative methods, such as Minres [68] or MINRES-QLP [13], which obtain very low data storage requirements. For the design of quasi Newton methods, its is crucial to note the implication that mainly symmetric quasi Newton methods should be of interest to us. Let us point out that in general both generalized derivatives operators ∂_B and ∂ are not linear, only

$$\partial(F_1 + F_2)(Z) \subset \partial F_1(Z) + \partial F_2(Z) \tag{62}$$

holds true. However for the generalized second derivative of our reduced Lagrangian we still obtain the linearity that we would expect for normal F-derivatives.

Lemma 2.23. *Let $Z \in E$ be arbitrary. We have the equalities of sets:*

$$\partial_B^2 f(Z) = \tfrac{1}{2}\mathcal{H} - \tfrac{1}{2}\partial_B |Z|_{\mathcal{K}} \quad \textit{and} \quad \partial^2 f(Z) = \tfrac{1}{2}\mathcal{H} - \tfrac{1}{2}\partial |Z|_{\mathcal{K}}$$

Proof. Both equalities follow from the fact that $Z \mapsto \mathcal{H}[Z]$ is continuously differentiable with derivative $\mathcal{H}$. ∎

Now the real unknown in all of our analysis are of course the sets $\partial_B|Z|_{\mathcal{K}}$ and $\partial|Z|_{\mathcal{K}}$. We will investigate further general and special properties of these sets later on.

2.4 A Conceptual Algorithm

We have now finished introducing the general analytical "toolbox" and like to propose a road map. Let to this end $\nabla f : E \to E$ be defined as in the previous sections, i. e.

$$\nabla f(Z) = \tfrac{1}{2}\mathcal{H}[Z] - \tfrac{1}{2}|Z|_{\mathcal{K}} + \mathcal{R}$$

and let us give a brief outlook to our algorithmic sections, where we consider various realizations of the following conceptual algorithm meant to approximate some $Z^{(*)} \in E$ satisfying $\nabla f(Z^{(*)}) = 0$.

Conceptual Algorithm 2.24. *Let $\mathcal{H} : E \to E$ and $\nabla f : E \to E$ be defined as in the previous section.*

1. *Input an initial point $Z^{(0)} \in E$, a maximal number of iterations $\kappa \in \mathbb{N} \cup \{\infty\}$ and a tolerance $\epsilon \geq 0$. Set $k := 0$.*

2. *Return $Z^{(k)}$ and stop if one of the criteria $\left\|\nabla f(Z^{(k)})\right\| < \epsilon$ or $k \geq \kappa$ is satisfied.*

3. *Choose a self adjoint linear operator $\mathcal{V}^{(k)} : E \to E$*

4. *Choose $\delta_k \geq 0$ and a direction $\Delta Z^{(k)} \in E$ such that*

$$\left\|\tfrac{1}{2}(\mathcal{H} - \mathcal{V}^{(k)})[\Delta Z^{(k)}] + \nabla f(Z^{(k)})\right\| \leq \delta_k \tag{63}$$

   is satisfied.

5. *Choose a step length* $t_k \in \mathbb{R}$ *and set*

$$Z^{(k+1)} := Z^{(k)} + t_k \Delta Z^{(k)} \tag{64}$$

6. *Set* $k := k + 1$ *and go to 2.*

This conceptual algorithm is kept rather vague and realizations may be trivial. Note however that each step is satisfiable independently of all prior steps (For example for all choices of $\mathcal{V}^{(k)}$ in step 3, one can find some $\Delta Z^{(k)}$ and $\delta_k \geq 0$ satisfying the condition in step 4.). For $\epsilon = 0$ and $\kappa = \infty$ it will therefore generate an infinite sequence $\{Z^{(k)}\}_{k \in \mathbb{N}} \subset E$. In order to transform the conceptual algorithm 2.24 into an implementable algorithm with $Z^{(k)} \overset{k \to \infty}{\longrightarrow} Z^{(*)}$, we obviously need to specify the choices of $\mathcal{V}^{(k)}$, δ_k, $\Delta Z^{(k)}$ and t_k. Whenever we choose $\delta_k = 0 \; \forall k \in \mathbb{N}_0$ a priori, we will call 2.24 exact and inexact otherwise. If we now want to design a well defined exact and globally convergent method from the conceptual algorithm 2.24, we have to choose $\mathcal{V}^{(k)}$ such that $\nabla f(Z^{(k)})$ is in the range of $\mathcal{H} - \mathcal{V}^{(k)}$ for each $k \in \mathbb{N}_0$. On the other hand we would also like $\mathcal{V}^{(k)}$ to contain as much (generalized) derivative information of $|.|_{\mathcal{K}}$ as possible in order to achieve fast convergence (which might contradict the prior rank condition and invertibility of $\mathcal{H} - \mathcal{V}^{(k)}$). In the following subsections we will therefore first consider (the) two extremal methods and afterwards present a "reasonable" compromise (a new norm constrained limited memory Quasi-Newton-Approach). Specifically, under the assumption of existence of some $Z^{(*)} \in E$ satisfying $\nabla f(Z^{(*)}) = 0$, we will investigate the following :

1. The choice of $\mathcal{V}^{(k)} = 0 \; \forall k \in \mathbb{N}_0$ leads to a first order or fixed point approach, which enjoys global (but potentially sub-linear) convergence and a low computational complexity per iteration. The exact worst-case convergence analysis leads us to a much broader concept, which is one of the main contributions in this thesis.

2. Choosing $\mathcal{V}^{(k)} \in \partial |Z^{(k)}|_{\mathcal{K}} \; \forall k \in \mathbb{N}_0$ results in a generalized Newton Approach, which is, under mild extra assumptions, locally (super-linearly) convergent, but suffers from high computational complexity per iteration.

3. Iteratively updated low rank matrices $\mathcal{V}^{(k)} = (\mathcal{P}^{(k)})^* \mathcal{W}^{(k)} \mathcal{P}^{(k)}$ containing partial (generalized) derivative information while also satisfying $||\mathcal{V}^{(k)}|| \leq 1 \; \forall k \in \mathbb{N}_0$ results in a novel norm constrained limited memory Quasi-Newton-Approach. It enjoys a moderate computational complexity per iteration and fast convergence in many practical examples.

3 Selected Euclidean Spaces and Cones

Before we finally jump into explicit algorithms we present a few relevant examples of Euclidean spaces and convex cones. We will use the following two subsections to briefly discuss first the real coordinate space and second the space of real symmetric matrices. In each of these two subsections we will then investigate the generalized absolute value function of some convex cones embedded in these Euclidean spaces. Before we begin however let us motivate working in the general setting of a Euclidean space equipped with some inner product. The reasoning for this non-specialized strategy is manifold. Let us start with two simple explanations.

The first reason is the ease of combining simpler spaces and cones: Let us assume that in (1) or (2) we are given the Cartesian products

$$E := E_1 \times \ldots \times E_s \qquad \text{and} \qquad \mathcal{K} := \mathcal{K}_1 \times \ldots \times \mathcal{K}_s \tag{65}$$

where E is equipped with the inner product $\langle .,. \rangle_E := \sum_{i=1}^{s} \langle .,. \rangle_{E_i}$ for a collection of Euclidean spaces E_i equipped with inner products $\langle .,. \rangle_{E_i}$ and an embedded collection of convex cones $\mathcal{K}_i \subset E_i$ for $1 \leq i \leq s$. Let us further assume that we also know how to (cheaply) evaluate the generalized absolute value functions $|.|_{\mathcal{K}_i} : E_i \to E_i$. Then we can evaluate the combined absolute value function $|.|_{\mathcal{K}} : E \to E$ for an argument $Z = (Z_i)_{i=1}^{s} \in E$ easily as

$$|Z|_{\mathcal{K}} = \left(|Z_i|_{\mathcal{K}_i}\right)_{i=1}^{s} \tag{66}$$

and therefore employ the same techniques we described in the previous sections.

The second reason for working in a general setting is the possibility to decompose an intersection of convex cones easily: Let us now assume that our convex cone $\mathcal{K}$ in (1) is given as

$$\mathcal{K} := \bigcap_{i=1}^{s} \mathcal{K}_i \tag{67}$$

for some convex cones $\mathcal{K}_i \subset E$. Now assume the evaluation of $|.|_{\mathcal{K}} : E \to E$ to be unavailable (or unreasonably expensive to compute), but the evaluation of each absolute value function $|.|_{\mathcal{K}_i} : E \to E$ for $1 \leq i \leq s$ to be easy enough. Then we can reformulate (1) as an equivalent problem of the same form over the convex cone $\tilde{\mathcal{K}} := \mathcal{K}_1 \times \ldots \times \mathcal{K}_s$ for example by exploiting the fact that $\mathcal{K} = \{X_1 \mid (X_i)_{i=1}^{s} \in \tilde{K},\ X_1 = X_j\ \forall\, 2 \leq j \leq s\}$. As described above, the cost of evaluating $|.|_{\tilde{K}} : E^s \to E^s$ is bounded above by the sum of cost of the evaluations of $|.|_{\mathcal{K}_i}$ for $1 \leq i \leq s$, which again enables us to use the techniques from the prior sections. We stretch the fact that great care is necessary when employing such equivalent reformulations, to in particular avoid redundant constraints within the new reformulation. We will address this topic again later on.

Let us conclude that the general Euclidean setting is flexible and expandable, without being overly complicated. What we can also conclude from this short motivation is the importance of 'simple' Euclidean spaces and 'simple' convex cones since they may give rise to the solution of problems with a much higher complexity.

3.1 Real Coordinate Space

In this subsection we will consider the Euclidean-Space $E = \mathbb{R}^n$ equipped with the standard Scalar-product, i.e. $\langle x, y\rangle := x^T y \ \forall x, y \in \mathbb{R}^n$. From equation (1), we recover the following, well known, primal standard form:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & x \in Ax = b \\ & x \in \mathcal{K} \end{aligned} \tag{68}$$

for some $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, and a matrix $A \in \mathbb{R}^{m\times n}$. Without loss of generality we will assume $rank(A) = m \leq n$ throughout this section. Note that for appropriate choice of $\mathcal{K}$ equation (68) includes the primal standard form of linear programming ($\mathcal{K} = \mathbb{R}^n_+$) as well as second order programming ($\mathcal{K} = \mathbb{L}_2$), both of which we will address below. Also note that $\mathbb{R}^n$ is in fact itself a convex cone, for which the polar cone takes the form $(\mathbb{R}^n)^P = \{0\}$.

3.1.1 Nonnegative Orthant

It is well known that the nonnegative orthant

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \mid x_i \geq 0 \ \forall \ 1 \ \leq i \leq n\} \tag{69}$$

is a self-dual convex cone and therefore its polar cone is given by

$$(\mathbb{R}^n_+)^P = -\mathbb{R}^n_+ =: \mathbb{R}^n_-. \tag{70}$$

Note that the nonnegative orthant gives rise to one of the simplest and probably most studied areas of conic programming, i.e. linear programming. One reason for calling it simple is the easy solution of its membership problem: In fact we even can easily derive closed form projection formulas

$$\begin{aligned} &\Pi_{\mathbb{R}^n_+}(z) = (\max(z_i, 0))_{i=1}^n, \quad \Pi_{\mathbb{R}^n_-}(z) = (\min(z_i, 0))_{i=1}^n \\ &|z|_{\mathbb{R}^n_+} = (|z_i|)_{i=1}^n = -|z|_{\mathbb{R}^n_-} \end{aligned} \tag{71}$$

which can be computed exactly with no more than n sign changes. While we are not going to investigate simplex- or interior point methods here, we emphasize the existence of very efficient software for solving linear programs such as [15], [26] or [61] based on these methods. Note however that generalizations to our setting, i.e. a closed convex, but possibly non-self-dual, cone (e.g. $\mathcal{K} = \tilde{\mathcal{K}} \times \mathbb{R}^n_+$), can be tricky or even impossible. Let us now consider Clarke's generalized derivative of $|.|_{\mathbb{R}^n_+}$ at a given point $z = (z_i)_{i=1}^n \in \mathbb{R}^n$. It is well known that the generalized derivative of the "normal" absolute value takes the form

$$\partial|z_i| = \begin{cases} -1 & \text{if } z_i < 0 \\ [-1, 1] & \text{if } z_i = 0 \\ 1 & \text{if } z_i > 0 \end{cases} \tag{72}$$

implying a diagonal structure for the multidimensional case

$$\partial|z|_{\mathbb{R}^n_+} = \text{Diag}((\partial|z_i|)_{i=1}^n) \tag{73}$$

which we can readily use to prove the following result regarding the existence of zeros of ∇f from the previous subsection.

Proposition 3.1. *Assume that* (68) *for* $\mathcal{K} = \mathbb{R}^n_+$ *has an optimal solution* $x^{(*)} \in \mathbb{R}^n_+$. *If we define* $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ *as before with* $E = \mathbb{R}^n$ *and* $\mathcal{K} = \mathbb{R}^n_+$, *i.e.*

$$\nabla f(z) = (I - A^T(AA^T)^{-1}A)[\Pi_{\mathbb{R}^n_-}(Z)) + c] - A^T(AA^T)^{-1}[A(\Pi_{\mathbb{R}^n_+}(z)) - b] \tag{74}$$

then there exists $z^{(*)} \in \mathbb{R}^n$ *such that* $\nabla f(z^{(*)}) = 0$. *If* $z^{(*)} \in \mathbb{R}^n$ *is in addition unique, i.e. the unique solution of* $\nabla f(z) = 0$ *then* $\partial^2 f(z^{(*)})$ *is single valued and its element is invertible.*

Proof. Since (68) has an optimal solution and is a linear program (for $\mathcal{K} = \mathbb{R}^n_+$) we know that its dual must also attain an optimal solution. We also know that these solutions will satisfy the conic KKT-conditions (4), which together with corollary 2.8. proves the existence of some $z^{(*)} \in \mathbb{R}^n$ such that $\nabla f(z^{(*)}) = 0$. If $z^{(*)} \in \mathbb{R}^n$ is the unique solution to the equation then $|z^{(*)}|_{\mathbb{R}^n_+}$ must be componentwise nonzero (since the solutions to both the primal and dual problem are unique, they must also be strictly complementary in the linear setting, see for example [93]). In other words $|.|_{\mathbb{R}^n_+}$ is continuously differentiable at $z^{(*)}$ and therefore ∇f is too. We conclude that $\partial^2 f(z^{(*)})$ is single valued. The rest of this proof follows via contradiction: Assume that the element in $\partial^2 f(z^{(*)})$ is not invertible, then there exists $0 \neq \delta z \in \mathbb{R}^n$ such that $\partial^2 f(z^{(*)})\delta z = 0$ and $sign(z) = sign(z + \delta z)$. We conclude that $\partial|z^{(*)}|_{\mathbb{R}^n_+} = \partial|z^{(*)} + \delta z|_{\mathbb{R}^n_+}$ and therefore (with a slight abuse of notation)

$$\begin{aligned}\nabla f(z^{(*)} + \delta z) &= \underbrace{\nabla f(z^{(*)})}_{=0} + \underbrace{\partial^2 f(z^{(*)})\delta z}_{=0} - \frac{1}{2}\Big(\underbrace{|z^{(*)} + \delta z|_{\mathbb{R}^n_+}}_{=\partial|z^{(*)}+\delta z|_{\mathbb{R}^n_+}(z+\delta z)} - |z^{(*)}|_{\mathbb{R}^n_+} - \partial|z^{(*)}|_{\mathbb{R}^n_+}\delta z\Big) \\ &= -\frac{1}{2}\Big(\partial|z^{(*)}|_{\mathbb{R}^n_+}(z^{(*)} + \delta z) - \underbrace{|z^{(*)}|_{\mathbb{R}^n_+}}_{=\partial|z^{(*)}|_{\mathbb{R}^n_+} z} - \partial|z^{(*)}|_{\mathbb{R}^n_+}\delta z\Big) \\ &= 0\end{aligned} \tag{75}$$

holds true, which directly contradicts the uniqueness of $z^{(*)} \in \mathbb{R}^n$. ■

As a consequence, if we are sure that there exists exactly one zero ∇f, then the standard Newton-method is locally (close to the unique zero) well defined. However, as reassuring the above theorem might be, a slight modification of its proof reveals the following truth.

Lemma 3.2. *Assume that* (68) *for* $\mathcal{K} = \mathbb{R}^n_+$ *has two distinct optimal solutions* $x^{(*1)}, x^{(*2)} \in \mathbb{R}^n_+$, $x^{(*1)} \neq x^{(*2)}$. *Then the equation* $\nabla f(z) = 0$, *for* $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ *defined as in* (74), *will have two distinct solutions* $z^{(*1)}, z^{(*2)} \in \mathbb{R}^n$, $z^{(*1)} \neq z^{(*2)}$ *and for every solution* $z^{(*)} \in \mathbb{R}^n$ *with* $\nabla f(z^{(*)}) = 0$ *there exists a singular element* $G \in \partial^2_B f(z^{(*)})$.

Proof. Again the existence of $z^{(*1)}, z^{(*2)} \in \mathbb{R}^n$, $z^{(*1)} \neq z^{(*2)}$ follows from strong duality for linear programs. Let us assume without loss of generality, that $z^{(*2)} \in \mathbb{R}^n$ is componentwise non-zero. Then there exists $1 \geq \epsilon > 0$ such that the points $(1-\lambda)z^{(*1)} + \lambda z^{(*2)} \in \mathbb{R}^n$ are also componentwise non-zero for all $\lambda \in (0, \epsilon)$. We conclude that $\nabla f(Z) = \frac{1}{2}\mathcal{H}[z] - \frac{1}{2}|z|_{\mathbb{R}^n_+} + \mathcal{R}$

must be continuously differentiable in all of these points. By possibly reducing $\epsilon > 0$ we can assume $\nabla^2 f((1-\lambda)z^{(*1)} + \lambda z^{(*2)})[z^{(*1)}] = \frac{1}{2}(\mathcal{H}[z^{(*1)}] - |z^{(*1)}|_{\mathbb{R}^n_+}) \ \forall \lambda \in (0, \epsilon)$

$$\begin{aligned}
0 &= \nabla f((1-\lambda)z^{(*1)} + \lambda z^{(*2)}) \\
&= \nabla^2 f((1-\lambda)z^{(*1)} + \lambda z^{(*2)})[(1-\lambda)z^{(*1)} + \lambda z^{(*2)}] + \mathcal{R} \\
&= \underbrace{\nabla^2 f((1-\lambda)z^{(*1)} + \lambda z^{(*2)})[z^{(*1)}] + \mathcal{R}}_{=\nabla f(z^{(*1)})=0} + \lambda \nabla^2 f((1-\lambda)z^{(*1)} + \lambda z^{(*2)})[z^{(*2)} - z^{(*1)}] \\
&= \underbrace{\lambda}_{>0} \nabla^2 f((1-\lambda)z^{(*1)} + \lambda z^{(*2)})[\underbrace{z^{(*2)} - z^{(*1)}}_{\neq 0}]
\end{aligned} \tag{76}$$

implying that $\nabla^2 f((1-\lambda)z^{(*1)} + \lambda z^{(*2)})$ is singular for all $\lambda \in (0, \epsilon)$. By taking the limit $\lambda \to 0$ we conclude that $G := \lim_{\lambda \to 0} \nabla^2 f((1-\lambda)z^{(*1)} + \lambda z^{(*2)}) \in \partial_B^2 f(z^{(*1)})$ exists and is a singular element of $\partial_B^2 f(z^{(*1)})$. ■

Now if we come back to the bigger picture, we see that assuming invertibility of all elements in the second generalized derivative of the reduced Lagrangian is a relatively strong assumption even for the "simple" cone $\mathbb{R}^n_+$ (which one should keep in mind for section 5). Changing topics back to to the general theme of this section let us next consider slightly more complex cones.

3.1.2 Weighted p-order Cones

Defining the componentwise multiplication of two vectors $x \circ y = (x_i y_i)_{i=1..n}$ for $x = (x_i)_{i=1..n}, y = (y_i)_{i=1..n} \in \mathbb{R}^n$ as well as the p norm $(1 \leq p)$ via

$$||x||_p := (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}} \tag{77}$$

for any $x = (x_i)_{i=1..n} \in \mathbb{R}^n$ gives rise to an infinite number of closed, convex cones. Namely for $\omega = (w_0, w) \in \mathbb{R} \times \mathbb{R}^n$ we define

$$\mathbb{L}_p^\omega := \{(x_0, x) \in \mathbb{R} \times \mathbb{R}^n \mid ||x \circ w||_p \leq x_0 w_0\} \subset \mathbb{R} \times \mathbb{R}^n \tag{78}$$

the so-called weighted p-order cones (sometimes referred to as weighted norm cones). If ω only contains ones we omit the term in the definition and simply refer to them as p-order cones, where the dimension will be evident from the context. One of the main benefits of these cones, is the possibility to reformulate many convex (and often non-differentiable) optimization problems in our standard format (1). For example

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad c^T x + ||Ax - b||_2 + ||x||_1 \tag{79}$$

which is often found in sparse recovery applications, can be equivalently stated in its epigraph formulation

$$\begin{aligned}
\underset{x \in \mathbb{R}^n, r \in \mathbb{R}^m, t_2, t_1 \in \mathbb{R}}{\text{minimize}} \quad & c^T x + t_2 + t_1 \\
\text{subject to} \quad & Ax - r = b \\
& ||r||_2 \leq t_2 \\
& ||x||_1 \leq t_1
\end{aligned} \tag{80}$$

and readily be seen to be of the form (1) for appropriate space and operator choices ($\mathcal{K} = \mathbb{L}_2 \times \mathbb{L}_1 \subset (\mathbb{R} \times \mathbb{R}^m) \times (\mathbb{R} \times \mathbb{R}^n) =: E$). Now while the second order cone $\mathbb{L}_2$ is self-dual (with respect to the scalar product $\langle x, y \rangle = \sum_{i=0}^{n} x_i y_i$) and the following closed form projection formula exists (from [2] 3.3.6)

$$\Pi_{\mathbb{L}_2}((z_0, z)) = \begin{cases} (z_0, z) & \text{if } ||z||_2 \leq z_0 \\ (0, 0) & \text{if } ||z||_2 \leq -z_0 \\ \frac{||z||_2 + z_0}{2}(1, \frac{z}{||z||_2}) & \text{else} \end{cases} \tag{81}$$

for $(z_0, z) \in \mathbb{R} \times \mathbb{R}^n$, this can not be said for general (weighted) p-order cones. Nevertheless we can often compute the projections quite efficiently anyway. For example we can project onto the weighted first and infinite order cones $\mathbb{L}_1^{\omega_1}$ and $\mathbb{L}_\infty^{\omega_\infty}$, by exploiting a certain duality. Let us be more precise here: For an arbitrary $z \in \mathbb{R} \times \mathbb{R}^n$ we are concerned with the computation of the Euclidean projection, i.e. the unique solutions $\Pi_{\mathbb{L}_1^\omega}(z)$ and $\Pi_{\mathbb{L}_\infty^\omega}(z)$ of

$$\text{minimize } \|z - x\|_2^2 \;\mid x \in \mathbb{L}_1^\omega \quad \text{respectively} \quad \text{minimize } \|z - x\|_2^2 \;\mid x \in \mathbb{L}_\infty^\omega \tag{82}$$

where we identified $\mathbb{R} \times \mathbb{R}^n \cong \mathbb{R}^{n+1}$ for matters of convenience. Note that the indexing starts at zero until the end of this subsection. Let us denote the pointwise division by $\frac{1}{\omega}$ with the convention that $w_j = 0$ implies that the j-th component of elements in $\mathbb{L}_\infty^{\frac{1}{\omega}}$ and $\mathbb{L}_1^{\frac{1}{\omega}}$ are both zero. Applying the definition of the dual cone $\mathcal{K}^D := \{y \in \mathbb{R}^{n+1} \mid y^T x \geq 0 \; \forall x \in \mathcal{K}\}$ and the polar cone $\mathcal{K}^P = -\mathcal{K}^D$, leads us to the following duality relation between $\mathbb{L}_1^{\omega_1}$ and $\mathbb{L}_\infty^{\omega_\infty}$.

Lemma 3.3. *For all $\omega \in \mathbb{R}_{\neq 0} \times \mathbb{R}^n$*

- $(\mathbb{L}_1^\omega)^D = \mathbb{L}_\infty^{\frac{1}{\omega}}$ *and* $(\mathbb{L}_\infty^\omega)^D = \mathbb{L}_1^{\frac{1}{\omega}}$
- $z = \Pi_{\mathbb{L}_1^\omega}(z) - \Pi_{\mathbb{L}_\infty^{\frac{1}{\omega}}}(-z)$

Note that the case $w_0 = 0$ is excluded for reasons of convenience in our notation.[5]

Proof. The first part follows immediately from the Hölder inequality. The second part is then simply Moreaus decomposition [60]. ■

I am not aware of any projection algorithm for the weighted cones $\mathbb{L}_1^\omega$ and $\mathbb{L}_\infty^\omega$ in the literature. However, in [75] an algorithm for projecting on the weighted l1-ball is discussed. The following is a slight modification of the ideas from [75] to fit the conic setting considered here. Since the conic projection algorithm is very similar to the l1-ball projection algorithm, the notation is kept as close as possible. Before we state an algorithm that computes $\Pi_{\mathcal{K}}(z)$ for any $\omega, z \in \mathbb{R}^n$, we will briefly describe the main ideas. We will not prove the algorithms integrity. First we need a few reduction steps. The previous Lemma allows us to solve either one of the problems in (82), so let $\mathcal{K} \in \{\mathbb{L}_1^\omega, \mathbb{L}_\infty^\omega\}$ for some $\omega \in \mathbb{R}_{\neq 0}$. There are few simple

[5] If $w_0 = 0$, $w_i \neq 0 \; \forall i \in \{1, .., n\}$ then $\Pi_{\mathbb{L}_1^\omega}(z) = \Pi_{\mathbb{L}_\infty^\omega}(z) = \begin{pmatrix} z_0 \\ 0 \end{pmatrix}$. In our notation we have $\mathbb{L}_1^\omega = \mathbb{R} \times \{0\}^n$ and $(\mathbb{L}_1^\omega)^P = \{0\} \times \mathbb{R}^n \neq -\mathbb{L}_\infty^{\frac{1}{\omega}}$.

observations to be made: First note that we can change signs of $w_j \ \forall j \in \{1,..,n\}$ without changing the solution of (82). Secondly we see that changing the sign of j-th component z_j for some $j \in \{1,..,n\}$ leads to a change of signs in the j-th component of the projection. Specifically we will use $\Pi_{\mathcal{K}}(\begin{pmatrix} z_0 \\ (z_i)_{i=1..n} \end{pmatrix} = \begin{pmatrix} 1 \\ (sign(z_i))_{i=1..n} \end{pmatrix} \circ \Pi_{\mathcal{K}} \begin{pmatrix} z_0 \\ (|z_i|)_{i=1..n} \end{pmatrix}$. Thirdly we will exploit symmetry, namely the fact that for any permutation $\tau : \{1,..,n\} \to \{1,..,n\}$ it holds that $(\Pi_{\mathcal{K}} \begin{pmatrix} z_0 \\ (z_{\tau(i)})_{i=1..n} \end{pmatrix})_j = (\Pi_{\mathcal{K}} \begin{pmatrix} z_0 \\ (z_i)_{i=1..n} \end{pmatrix})_{\tau(j)} \ \forall j \in \{1,..,n\}$. Last but not least if the j-th component of z is equal to zero for some $j \in \{1,..,n\}$ then so will the j-th component of its projection ($\exists j \in \{1..n\} : z_j = 0 \Rightarrow (\Pi_{\mathcal{K}} \begin{pmatrix} z_0 \\ (z_i)_{i=1..n} \end{pmatrix})_j = 0$). We are now in the situation that we can assume without loss of generality $\omega = \begin{pmatrix} w_0 \\ w \end{pmatrix} \in \mathbb{R}_{\neq 0} \times \mathbb{R}^n_{++}$, $z = \begin{pmatrix} z_0 \\ (z)_{i=1,..,n} \end{pmatrix} \in \mathbb{R} \times \mathbb{R}^n_{++}$ such that $\frac{z_1}{w_1} \geq \frac{z_2}{w_2} \geq ... \geq \frac{z_n}{w_n} > 0$. It is then easy to see that the following algorithm computes for any $z, \omega \in \mathbb{R}^{n+1}$ the projection $\Pi_{\mathbb{L}_1^\omega}(z)$. We omit the details.

Algorithm 3.4. *Projection of $z \in \mathbb{R}^{n+1}$ on $\mathbb{L}_1^\omega(z)$*

0 Input: *A weighting vector $\omega = \begin{pmatrix} w_0 \\ (w_i)_{i=1..n} \end{pmatrix} \in \mathbb{R}^{n+1}$ and an arbitrary $z = \begin{pmatrix} z_0 \\ (z_i)_{i=1..n} \end{pmatrix} \in \mathbb{R}^{n+1}$.*

1 Preprocessing:

(a) Set $S := \{i \in \{1..n\} \mid w_i \neq 0\} \cap \{i \in \{1..n\} \mid z_i \neq 0\}$

i. If $w_0 = 0$ set $(\Pi_{\mathbb{L}_1^\omega}(z))_i := z_i \ \forall i \notin S$, $(\Pi_{\mathbb{L}_1^\omega}(z))_i := 0 \ \forall i \in S$ and return

ii. If $|S| = 0$ and $w_0 z_0 < 0$ set $\Pi_{\mathbb{L}_1^\omega}(z) := \begin{pmatrix} 0 \\ (z_i)_{i=1..n} \end{pmatrix}$ and return

iii. If $|S| = 0$ and $w_0 z_0 \geq 0$ set $\Pi_{\mathbb{L}_1^\omega}(z) := z$ and return

iv. Set $(\Pi_{\mathbb{L}_1^\omega}(z))_i := z_i \ \forall i \notin S$

(b) If $\sum_{i \in S} |w_i z_i| \leq w_0 z_0$ set $(\Pi_{\mathbb{L}_1^\omega}(z))_i := z_i \ \forall i \in S$ and return

(c) If $\max_{i \in S} |\frac{z_i}{w_i}| \leq -\frac{z_0}{w_0}$ set $(\Pi_{\mathbb{L}_1^\omega}(z))_i := 0 \ \forall i \in S \cup \{0\}$ and return

2 Computation:

(a) Find Permutation: $\tau : \{1,..,|S|\} \to S$ such that $|\frac{z_{\tau(1)}}{w_{\tau(1)}}| \geq |\frac{z_{\tau(2)}}{w_{\tau(2)}}| \geq ... \geq |\frac{z_{\tau(|S|)}}{w_{\tau(|S|)}}| > 0$

(b1) Find the smallest $k^* \in \{0,..,|S|-1\}$ such that

$$|\frac{z_{\tau(|S|-k^*)}}{w_{\tau(|S|-k^*)}}| > \frac{\sum_{i=1,..,|S|-k^*} |w_{\tau(i)} z_{\tau(i)}| - w_0 z_0}{\sum_{i=0,...,|S|-k^*} w^2_{\tau(i)}} \tag{83}$$

(b2) or find the largest $j^* \in \{1,..,|S|\}$ such that

$$|\frac{z_{\tau(j^*)}}{w_{\tau(j^*)}}| > \frac{\sum_{i=1,..,j^*} |w_{\tau(i)} z_{\tau(i)}| - w_0 z_0}{\sum_{i=0,...,j^*} w^2_{\tau(i)}} \tag{84}$$

and set $k^* = |S| - j^* + 1$

3 Output: *Set*

$$
\begin{aligned}
(\Pi_{\mathbb{L}_1^\omega}(z))_0 &= -(-z_0 - w_0 \frac{\sum_{i=1,..,|S|-k^*} |w_{\tau(i)} z_{\tau(i)}| - w_0 z_0}{\sum_{i=0,...,|S|-k^*} w_{\tau(i)}^2}) \\
(\Pi_{\mathbb{L}_1^\omega}(z))_{\tau(j)} &= sign(z_{\tau(j)})(|z_{\tau(j)}| - |w_{\tau(j)}| \frac{\sum_{i=1,..,|S|-k^*} |w_{\tau(i)} z_{\tau(i)}| - w_0 z_0}{\sum_{i=0,...,|S|-k^*} w_{\tau(i)}^2}) \ \forall j \in \{1,..,|S|-k^*\} \\
(\Pi_{\mathbb{L}_1^\omega}(z))_{\tau(j)} &= 0 \ \forall j \in \{|S|-k^*+1,..,|S|\}
\end{aligned}
\tag{85}
$$

and return

Note that the above algorithm can be implemented with no more than $\mathcal{O}(nlog_2(n))$ sorting and $\mathcal{O}(n)$ arithmetic operations. For large $n, |S| \in \mathbb{N}$ the cost for finding the permutation τ will generally dominate the cost of all other operations.

3.2 Space of Real Symmetric Matrices

In this subsection we will consider the Euclidean space $E = \mathbb{S}^n$ where

$$\mathbb{S}^n := \{Z \in \mathbb{R}^{n \times n} \mid Z = Z^T\} \tag{86}$$

is the space of real symmetric matrices equipped with the trace product $\langle X, Y \rangle := X \bullet Y \ \forall X, Y \in \mathbb{S}^n$ where $X \bullet Y := \text{trace}(X^T Y) \ \forall X, Y \in \mathbb{R}^{n \times n}$. Note that the trace product, defined above, is actually a scalar product not only on the space of symmetric matrices, but also on the space of rectangular real matrices $\mathbb{R}^{n \times k}$. This explains the use of the otherwise superfluous transpose symbol within. The norm induced by the trace product is widely known as the Frobenius-Norm and usually denoted by $||Z||_F^2 := \text{trace}\,(Z^T Z)$.

3.2.1 Semidefinite Cone

The probably most famous representative of convex cones in $\mathbb{S}^n$ is given as

$$\mathbb{S}^n_+ := \{X \in \mathbb{S}^n \mid y^T X y \geq 0 \ \forall \ y \in \mathbb{R}^n\} \tag{87}$$

and commonly referred to as positive semidefinite cone (or sometimes just semidefinite cone). It is closed, convex, nonempty and self-dual (with respect to $\bullet$). We write for its polar cone $(\mathbb{S}^n_+)^P = -\mathbb{S}^n_+ =: \mathbb{S}^n_-$. Let us, in order to project some $Z \in \mathbb{S}^n$ onto the positive semidefinite cone, assume that a spectral decomposition has been computed a priori, i.e.

$$Z = V \,\text{Diag}(d) V^T \text{ with } V \in \mathbb{R}^{n \times n} \text{ such that } VV^T = I_n \text{ and } d \in \mathbb{R}^n \tag{88}$$

are available. In terms of this decomposition, we obtain the following well known identities

$$\Pi_{\mathbb{S}^n_+}(Z) = V \,\text{Diag}(\max(d_i, 0)_{i=1}^n) V^T, \tag{89}$$

$$\Pi_{\mathbb{S}^n_-}(Z) = V \,\text{Diag}(\min(d_i, 0)_{i=1}^n) V^T \tag{90}$$

and

$$|Z|_{\mathbb{S}^n_+} = V \,\text{Diag}(|d_i|_{i=1}^n) V^T \tag{91}$$

for which we address the computational evaluation cost: Let us first note that, in general, finding a spectral decomposition can not be done in an exact fashion[6] (unless $n \in \mathbb{N}$ is sufficiently small or $Z \in \mathbb{S}^n$ has some additional structure). However efficient iterative methods exist and an approximate spectral decomposition of Z can for example be realized by employing QR or Divide-and-conquer eigenvalue algorithms. Both of which usually need $\mathcal{O}(n^3)$ arithmetic operations to reach satisfactory precision. While it is difficult to choose from the wide variety of possible eigenvalue algorithms, the Divide-and-conquer eigenvalue algorithm employed by the LAPACK-Routine dsyevd in our numerical experiments is a convincing choice, in terms of both speed and accuracy, for the setting of a single CPU with multiple cores. In the setting of a heterogeneous computer architecture, for example with one CPU and two GPUs, its hybrid brother (it employs the CPU and both GPUs simultaneously), the MAGMA-Routine [87] magma_dsyevd_m (from version 1.6.2, in combination with Intels Math Kernel Library 2015), represent an even faster choice for sufficiently large $n \in \mathbb{N}$. To give an impression about the computational time needed, there is a MATLAB graphic included below. Note that the data used to generate the plot is by no means reproducible, mainly due to software and hardware upgrades, but also due to missing data points (actually two data points are interpolated, due to minor inconveniences on the High Performance Cluster). Note that all routines are either employed via Mex files or built-in MATLAB algorithms (the rather technical and not fully optimized details are omitted here). The point to be seen here is that we can, for large blocksizes (the largest matrix considered here is 15000×15000) greatly benefit, in terms of time, from new and/or advanced software as well as hardware.

[6]Note that this does not necessarily imply that $|Z|_{\mathbb{S}^n_+}$ can not be evaluated in an exact fashion: Its evaluation could still be possible via means that do not employ eigenvalues after all. For example note the equality $|Z|^2_{\mathbb{S}^n_+} = Z^2$ and the fact that we can easily find a non-symmetric square root of a positive definite matrix (namely through its Cholesky decomposition). While it remains unknown whether finding its (unique) positive semidefinite symmetric square root in an exact fashion is possible, there exist methods to approximate it iteratively (see for example [62] and [30]).

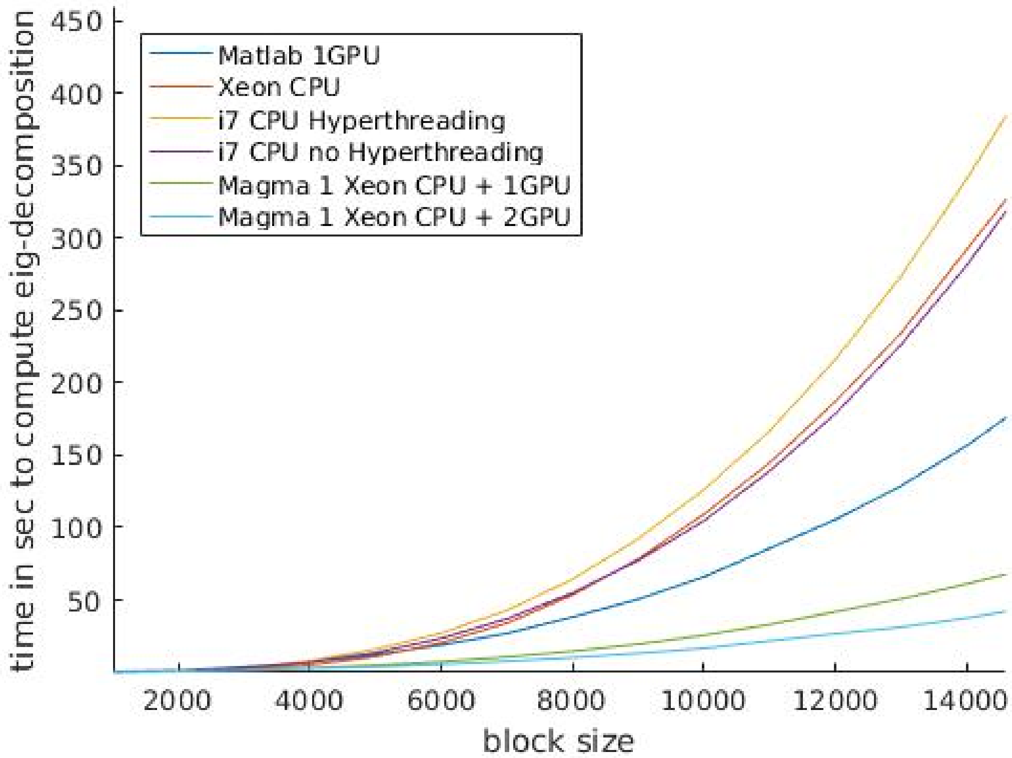


Figure 1: Approximate averaged timing in seconds to compute one symmetric eigenvalue decompositions via divide and conquer (not reproducible and meant only to give a very rough idea). We used a desktop pc with an Intel desktop CPU (Core(TM) i7-4770 CPU @ 3.40GHz), as well as HPC-node(s) with a server CPU (Xeon) together with NVIDIA Tesla K20Xm GPUs. Due to non-existing administrative privileges on the HPC node, we unfortunately could not fully embrace the power of the Xeon CPU.

Now let us focus on the remaining computational cost to evaluate $|Z|_{S_+}$, i.e. the multiplication of a diagonal matrix with V and V^T: Let us assume without loss of generality that the vector $d \in \mathbb{R}^n$ is sorted such that $d = (d_-^T, d_0^T, d_+^T)^T$ with component wise strictly positive $d_+ \in \mathbb{R}^{n_+}_{++}$, a zero vector $d_0 = 0 \in \mathbb{R}^{n_0}$ and a component wise strictly negative vector $d_- \in \mathbb{R}^{n_-}_{--}$. Let us also partition $V = [V_-, V_0, V_+]$ analogously, i.e. $V_- \in \mathbb{R}^{n \times n_-}$, $V_0 \in \mathbb{R}^{n \times n_0}$ and $V_+ \in \mathbb{R}^{n \times n_+}$. Basic linear algebra shows that the projections now take the form

$$\Pi_{\mathbb{S}^n_+}(Z) = V_+ \operatorname{Diag}(d_+) V_+^T = (V_+ \operatorname{Diag}(d_+)^{\frac{1}{2}})(V_+ \operatorname{Diag}(d_+)^{\frac{1}{2}})^T \tag{92}$$

and

$$\Pi_{\mathbb{S}^n_-}(Z) = V_- \operatorname{Diag}(d_-) V_-^T = -(V_- \operatorname{Diag}(-d_-)^{\frac{1}{2}})(V_- \operatorname{Diag}(-d_-)^{\frac{1}{2}})^T. \tag{93}$$

Since we can derive $|Z|_{\mathbb{S}^n_+}$ from either one of the two formulas above via (14), we can conclude that this second part of its evaluation does not exceed the number of arithmetic operations of a symmetric rank k matrix multiplication[7] for $k := min(n_-, n_+) \leq \frac{n}{2}$ by more than $\mathcal{O}(n^2)$. This argument is quite compelling whenever k is much smaller then $\frac{n}{2}$. In fact it also reveals that it is not even necessary to compute a full spectral decomposition in the first place. If one has prior knowledge about the numbers n_+, n_0, n_- one could for example be tempted to only compute all negative or only all positive eigenvalues together with their eigenvectors. The reason for not following this temptation, is connected to the cheap evaluation of the derivative of the generalized absolute value: Let us start by introducing the well known (bijective) linear mapping $\text{svec} : \mathbb{S}^n \to \mathbb{R}^{\frac{n^2+n}{2}}$ defined via

$$\text{svec}\,(Z) := \left(Z_{11}\ \sqrt{2}Z_{21} \ldots \sqrt{2}Z_{n1},\ Z_{22}\ \sqrt{2}Z_{32} \ldots \sqrt{2}Z_{n2}, \ldots, Z_{(n-1)(n-1)}\ \sqrt{2}Z_{n(n-1)},\ Z_{nn}\right)^T \tag{94}$$

which is obtained by multiplying all off-diagonal entries of Z by $\sqrt{2}$ and then stacking all columns of the lower triangular part on top of each other. We can easily verify that $X \bullet Y = \text{svec}(X)^T \,\text{svec}(Y)$ holds for all $X, Y \in \mathbb{S}^n$. In other words, the mapping svec is an isometric isomorphism. We denote its inverse mapping $\text{smat} : \mathbb{R}^{\frac{n^2+n}{2}} \to \mathbb{S}^n$. The generalized absolute value $|.|_{\mathbb{S}^n_+}$ is (continuously) differentiable at all full rank matrices $Z \in \mathbb{S}^n$ (i.e. when d_0 is of dimension zero). However as its derivative is an operator acting on matrices, rather than vectors, it is matrix representable, but not truly a matrix itself. We can express the action (see for example [99]) via

$$\nabla |Z|_{\mathbb{S}_+}[\Delta Z] = \text{smat}((I \otimes_{\mathbb{S}} |Z|_{\mathbb{S}_+})^{-1}(I \otimes_{\mathbb{S}} Z)\,\text{svec}(\Delta Z)) \tag{95}$$

where the symmetrized Kronecker product $X \otimes_S Y$ is implicitly defined via

$$(X \otimes_{\mathbb{S}} Y)\,\text{svec}(S) := \tfrac{1}{2}\,\text{svec}(YSX^T + XSY^T) \tag{96}$$

for arbitrary $X, Y \in \mathbb{R}^{n \times n}$ and symmetric $S \in \mathbb{S}^n$. We refer to the Appendix of [85], which yields an excellent overview regarding properties of the symmetrized Kronecker product. The properties considered, especially yield

$$\begin{aligned} \nabla |Z|_{\mathbb{S}_+} &\cong (I \otimes_{\mathbb{S}} |Z|_{\mathbb{S}_+})^{-1}(I \otimes_{\mathbb{S}} Z) \\ &= (V \otimes_{\mathbb{S}} V)(I \otimes_{\mathbb{S}} \text{Diag}(|d|_{\mathbb{R}^n_+}))^{-1}(I \otimes_{\mathbb{S}} \text{Diag}(d))(V^T \otimes_{\mathbb{S}} V^T) \end{aligned} \tag{97}$$

for the eigenvalue decomposition $Z = V\,\text{Diag}(d)V^T$, which leads to

$$\nabla |Z|_{\mathbb{S}_+}[\Delta Z] = V(\Omega \circ (V^T \Delta Z V))V^T \tag{98}$$

for $\Omega := \frac{d_i + d_j}{|d_i| + |d_j|} \in \mathbb{S}^n$ for $i, j \in \{1, .., n\}$. By realizing that Ω has large blocks containing only ones or minus ones, this formula has been refined in [99], allowing relatively cheap evaluations of $\nabla |Z|_{\mathbb{S}_+}[\Delta Z]$. We will address possible computational benefits later on. For now let us shift our focus to the generalized second derivative of the reduced Lagrangian. When it comes to semidefinite programing, one of the major assumptions for showing fast convergence of interior point methods, is the existence of strictly complementary optimal

[7] In the presence of Strassen and Coppersmith-Winograd algorithms it is quite unlikely for a tight upper bound on the arithmetic operations, needed for a symmetric rank k matrix multiplication, to exist. A naive bound is obviously given by $\frac{(2k-1)n(n+1)}{2}$ additions and multiplications.

primal and dual solutions, i.e. that some optimal solution $X^{(*)} \in \mathbb{S}^n_+$ of (1) and an optimal solution $Y^{(*)} \in \mathbb{S}_-$ of (2) satisfy $X^{(*)} - Y^{(*)} \succ 0$. If we translate this assumption to our case, we realize that it implies that the generalized absolute value is continuously differentiable at $Z^{(*)} := X^{(*)} + Y^{(*)} \in \mathbb{S}^n$. In other words, the generalized second derivative of the reduced Lagrangian will be single valued and the element will be equal to the (usual) second F-derivative. Let us also point out, that by assuming (in addition to $|Z^{(*)}|_{\mathcal{K}} = X^{(*)} - Y^{(*)} \succ 0$) that $Z^{(*)}$ is not only one, but the unique zero of the gradient of the reduced Lagrangian we can even show invertibility of this second derivative. The proof of this fact is omitted; it would essentially follow the same reasoning, that was presented in [25]. This connection is however worth keeping in mind when it comes to second order methods, presented later on.

3.2.2 Completely Positive and Copositive Cone Relaxations

The last example is connected to the completely positive cone

$$\mathcal{C}_n^* := \operatorname{conv}\{xx^T \mid x \in \mathbb{R}^n_+\} \tag{99}$$

and its dual

$$\mathcal{C}_n := \{S \in \mathbb{S}^n \mid x^T S x \geq 0 \ \forall x \in \mathbb{R}^n_+\} \tag{100}$$

the copositive cone. These cones are of great interest, because they can be used to reformulate many NP-hard problems as conic optimization problems of the form (1). Let us recall such a reformulation for the Max-Stable-Set (respectively the Max-Clique problem). For a given simple graph $G = (V, E(G))$ with $n = |V|$ vertices V, we denote the simple complementary graph with $\bar{G} = (V, E(\bar{G}))$ and its adjacency matrix with A_G (respectively $A_{\bar{G}}$). Note that a simple graph is loopless and undirected. Therefore both A_G and $A_{\bar{G}}$ are symmetric and we get $A_G + A_{\bar{G}} + I = J$, where I denotes the identity and J the matrix of all ones in $\mathbb{R}^{n\times n}$. In order to keep our notation short, we will define two vectors $X_{A_G} := (\frac{X_{ij}+X_{ji}}{\sqrt{2}})_{\{i,j\}\in E(G)} \in \mathbb{R}^{|E(G)|}$ and $X_{A_{\bar{G}}} := (\frac{X_{ij}+X_{ji}}{\sqrt{2}})_{\{i,j\}\in E(\bar{G})} \in \mathbb{R}^{|E(\bar{G})|}$ where $|E(G)|$ and $|E(\bar{G})|$ denote the number of edges in G and $\bar{G}$. The Max-Stable-Set problem can be stated as follows: Find a maximal subgraph $G' \subset G$ such that the complementary graph $\bar{G}'$ is a complete graph. The stability number $\alpha(G)$ is then defined as the number of vertices of G'. It was shown in [23] (with different scaling) that the following equivalent conic formulation

$$\begin{aligned} \alpha(G) = \max_{X\in\mathbb{S}^n} \quad & J \bullet X \\ \text{subject to} \quad & X_{A_G} = 0, \ \frac{I \bullet X}{\sqrt{n}} = \frac{1}{\sqrt{n}} \\ & X \in \mathcal{C}_n^* \end{aligned} \tag{101}$$

and its dual problem both attain optimal solutions with a non-zero duality gap. In theory we could therefore use any of the algorithms (that we are going to present next) to approximate the stability number. However, since tracing $\mathcal{C}_n^*$ is in general an NP-hard problem, it is unlikely that the projection operator $\Pi_{\mathcal{C}^*}$ can be computed efficiently in practice. If one is instead interested in finding upper bounds on the max clique number we can replace the completely positive cone in (101) by a "simpler" outer approximation. For example the doubly nonnnegative cone

$$\mathcal{DNN} := \mathbb{S}^n_+ \cap \mathbb{R}^{n\times n}_+ \tag{102}$$

for which the dual cone is given by the nonnegatively decomposable cone

$$\mathcal{NND} := (\mathcal{DNN})^D = \mathbb{S}^n_+ + \mathbb{R}^{n\times n}_+ \tag{103}$$

is a suitable choice. After eliminating redundant constraints and some minor reformulations, the resulting problem can be stated as

$$\begin{aligned}
\theta^+(G) = \max_{X\in\mathbb{S}^n, x_{\bar{G}}\in\mathbb{R}^{|E(\bar{G})|}} \quad & J \bullet X \\
\text{subject to} \quad & X_{A_G} = 0,\ \frac{I \bullet X}{\sqrt{n}} = \frac{1}{\sqrt{n}} \\
& \frac{X_{A_{\bar{G}}} - x_{\bar{G}}}{\sqrt{2}} = 0,\ x_{\bar{G}} \in \mathbb{R}^{|E(\bar{G})|}_+ \\
& X \in \mathbb{S}^n_+
\end{aligned} \tag{104}$$

and its optimal value $\theta^+(G)$ is usually refered to as Lovasz-Schrijver-number. Relaxing the problem even further leads to the so-called Lovasz-number which is defined as the following optimal value

$$\begin{aligned}
\theta(G) := \max_{X\in\mathbb{S}^n} \quad & J \bullet X \\
\text{subject to} \quad & X_{A_G} = 0,\ \frac{I \bullet X}{\sqrt{n}} = \frac{1}{\sqrt{n}} \\
& X \in \mathbb{S}^n_+
\end{aligned} \tag{105}$$

and because of the inclusions

$$\mathcal{C}^* \subset \mathcal{DNN} \subset \mathbb{S}^n_+ \tag{106}$$

we obtain the well known inequalities

$$\alpha(G) \leq \theta^+(G) \leq \theta(G). \tag{107}$$

Note that the affine constraints in (101),(104) and (105) are scaled such that the associated linear operators have orthonormal (with respect to the canonical scalar product) "rows", i.e. the resulting linear operator(s) $\mathcal{A}\mathcal{A}^*$ will be equal to some identity operator of appropriate size.

4 First Order Approach

In the following subsections we will loosen our setting (to general nonexpansive operators in Hilbert spaces). As a motivation, we first consider the conceptual algorithm 2.24 with $\mathcal{V}^{(k)} = 0\ \forall k \in \mathbb{N}$. One may speak of a first order approach, because the operator of the linear system is fixed and every (variable) curvature information of $\langle |.|_{\mathcal{K}}, .\rangle$ is disregarded. If we also choose $\delta_k = 0\ \forall k \in \mathbb{N}$, i.e. solve the linear system for $\Delta Z^{(k)} \in E$ exactly, we end up with the directions

$$\Delta Z^{(k)} = -2\mathcal{H}[\nabla f(Z^{(k)})] \tag{108}$$

and the simple iteration

$$Z^{(k+1)} = Z^{(k)} + t_k \Delta Z^{(k)} \tag{109}$$

which we will investigate in this and the next subsection. It turns our that iteration (109) is a version of Krasnoselski-Mann iteration ([43] and [57]), an observation that will be discussed later. Let us first start with a numerical consideration and a chronological ranging. In the early days of this thesis I implemented an inexact version of above iteration (i.e. small but positive $\delta_k > 0$, $\Delta Z^{(k)} \approx -2\mathcal{H}[\nabla f(Z^{(k)})]$ and fixed $t_k = t \in (0,1)\ \forall k \in \mathbb{N}_0$) in MATLAB (2014a) for the case $E = \mathbb{S}^n$ and $\mathcal{K} = \mathbb{S}^n_+$. This first[8] implementation behaved quite pleasing, especially its comparison to other methods (see details below). After more numerical testing and validation, I intended to explain the observed behavior mathematically. I almost instantly hit a brick wall: Although proving (monotone) convergence is an easy task, the obtained rates of convergence remained rather poor and in contrast to their empirical counterpart. Initially I was convinced to be overlooking some structure of ∇f coming from (1). As it will turn out in this section, this was not the case. After a short numerical consideration, we will start by recalling some of the known results regarding iteration (109), in particular one complexity bound. We will proceed by giving two examples of the form (1), proving that for fixed step length t the rather poor known upper complexity bounds on the convergence rate can not be improved significantly (implying that I did not overlook any structure before). We will complete this subsection with a result showing that these two examples actually reflect worst-case scenarios. The result is stated as a corollary of a theorem of the next subsection. This theorem is stated for finding fixed points, rather than finding zeros, since this seems more natural. Its proof and the resulting extensions, are some of the main contributions of this thesis. Specifically the technique used consists of numerical, statistical and algebraic tools to analyze complexities of given fixed point iterations in an optimal fashion. Let us get back to the promised numerical consideration and reveal some mathematically trivial, but numerically essential details. To start with, our input data $b \in \mathbb{R}^m$, $C \in \mathbb{S}^n$ is internally normalized in the sense that

$$||(\mathcal{I} - \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})[C]||_{\mathbb{S}^n} \approx ||\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}b||_{\mathbb{S}^n} \approx \frac{1}{\sqrt{2}} \tag{110}$$

should be approximately satisfied which implies

$$||\mathcal{R}||_{\mathbb{S}^n} \approx 1 \tag{111}$$

[8]The final (improved) implementation differs quite a lot, and we will therefore consider similar numerical tests later on for the improved version.

as well a certain primal and dual balancing. In each iteration we have to approximately solve one (not two) linear system of the form $\mathcal{A}\mathcal{A}^*y = r$, for which MATLABs standard preconditioned CG-Method (pcg) is used. An incomplete Cholesky factor[9] of $\mathcal{A}\mathcal{A}^*$ is computed as a preconditioner and the final iterate of the previous CG-iteration is then used as a starting point for the current CG-iteration. One finds that a relatively small tolerance is needed in order to solve all problems below, so we set $tol_k^{CG} := max[10^{-14}, 10^{-7}||\nabla f(Z^{(k)})||_{\mathbb{S}^n}]$. To approximate $|Z|_{\mathbb{S}^n_+}$ we can either use the same eigendecomposition that was used in [99] based on a divide and conquer strategy[10] or an inexact update strategy exploiting

$$|Z^{(k+j)}|_{\mathbb{S}^n_+} \approx |Z^{(k)}|_{\mathbb{S}^n_+} + \partial|Z^{(k)}|_{\mathbb{S}^n_+}[Z^{(k+j)} - Z^{(k)}] \tag{112}$$

whenever $||(Z^{(k+j)} - Z^{(k)})||_{\mathbb{S}^n} \leq \lambda_{min}(|Z^{(k)}|_{\mathbb{S}^n_+})$ and $1 \leq j \leq 50$, where $\lambda_{min}(.)$ denotes the smallest eigenvalue of its argument. Note that $\partial|Z|_{\mathbb{S}^n_+}[\Delta Z]$ can be computed with only $O(\gamma_k n^2)$ flops (see [99]), for $\gamma_k = \min[\#\{(\lambda(Z^{(k)})_i < 0)_{i=1}^n\}, \#\{(\lambda(Z^{(k)})_i > 0))_{i=1}^n\}]$ where $\lambda(.)$ denotes the vector of eigenvalues of its argument. These updates are computationally cheaper than a full eigenvalue decomposition and therefore counted (**up**). Infeasibility and optimality are measured (only once) a posteriori as follows:

$$R_D := \frac{||C + S^{(*)} - \mathcal{A}(y^{(*)})||_{\mathbb{S}^n}}{1 + ||C||_{\mathbb{S}^n}}, \quad R_P := \frac{||b - A(X^{(*)})||_2}{1 + ||b||_2}, \quad gap := \frac{b^T y^{(*)} - C \bullet X^{(*)}}{1 + |b^T y^{(*)}| + |C \bullet X^{(*)}|} \tag{113}$$

where $X^{(*)}, S^{(*)} := -Y^{(*)}$ and $y^{(*)}$ are computed from the final iterate $Z^{(k)}$ according to corollary 2.9 while also undoing our initial normalization. Note that we internally only rely on our (normalized) stopping criterion $||\nabla f(Z)||_{S^n} \leq tol$. **time** (in seconds), CPU-time (**cpu**), the number of iterations (**it**) are given as well as the total number of CG-iterations **CGit**. Note that time includes the time for computing an incomplete Cholesky factor.
It should be mentioned that the implementation allows warm-starts, i.e. benefits from "good" intial guesses $Z^{(0)} \in \mathbb{S}^n$. However in order to ensure comparability, we have restrained to $Z^{(0)} = 0 \in \mathbb{S}^n$ in all cases considered here. The tests here were performed on a Laptop with an Intel i7 quadcore 2,2Ghz processor and 8Gb memory running Ubuntu 14.04 and MATLAB 2014a. First, the implementation was tested on 24 random sparse problems considered in [56]. The generator for these problems can be found on Franz Rendls Webpage [74]. The random number generator is initialized with the parameter **seed**, where p controls density. n is the size of the SDP-Block, m is the number of constraints. The results were compared to the boundary point method [69] also available on Franz Rendls Webpage [74]. The boundary method stops when $max[R_P, R_D] \leq tol$. Let us note that for these problems, the boundary point method outperformed SDPNAL (see [99]). The numerical results of SDPNAL's performance are included below only for sake of completeness: Note that this does not properly reflect SDPNAL average performance (which is much better). The first test was performed with $tol = 10^{-6}$. The results are given in Table 1 and in Table 2 below:

[9] To be precise: An incomplete Cholesky factorization with threshold dropping (ICT) and drop tolerance 10^{-5}. (MATLAB-code: ichol(M, struct('type','ict','droptol',10−5))

[10] LAPACK's dsyevd. Let us however note, that the computational advantage over MATALB's eig routine is non-existent in MATLAB 2014a and newer, while it is up to 5 times faster in older versions.

seed	$(n^2+n)/2$	m	p	it	CGit	up	$C \bullet X$	$b^T y$	R_P	R_D	gap	time	cpu
3002030	45150	20000	3	100	213	46	7.61329e+02	7.61352e+02	1.7e-05	9.6e-08	1.5e-05	3.2	12.5
3002530	45150	25000	3	300	617	144	7.38372e+01	7.38384e+01	2.8e-06	1.5e-09	8.0e-06	15.9	60.2
3001040	45150	10000	4	166	341	65	1.65975e+02	1.65975e+02	4.7e-07	1.4e-08	1.4e-07	7.9	28.2
4003030	80200	30000	3	109	143	41	1.07214e+03	1.07214e+03	3.1e-06	5.5e-07	1.6e-06	4.6	19.4
4004030	80200	40000	3	167	350	59	8.05770e+02	8.05769e+02	2.8e-06	2.5e-09	1.1e-06	12.9	49.7
4001540	80200	15000	4	209	411	45	-6.55000e+02	-6.55000e+02	2.0e-07	7.5e-09	6.3e-08	16.1	64.4
5003030	125250	30000	3	158	169	58	1.10763e+03	1.10763e+03	6.8e-07	2.2e-08	1.1e-07	10.9	46.5
5004030	125250	40000	3	116	129	38	8.16611e+02	8.16611e+02	1.2e-06	1.5e-07	1.5e-07	9.6	40.1
5005030	125250	50000	3	103	140	27	3.64945e+02	3.64945e+02	1.5e-06	1.2e-07	2.7e-07	10.1	41.4
5002040	125250	20000	4	274	531	117	3.28004e+02	3.28004e+02	2.0e-07	9.2e-09	1.5e-08	22.3	86.2
6004030	180300	40000	3	171	182	79	3.06618e+02	3.06618e+02	1.2e-06	4.1e-08	6.1e-07	18.4	40.5
6005030	180300	50000	3	131	143	45	-3.86414e+02	-3.86413e+02	6.2e-07	1.5e-08	1.4e-06	17.5	38.1
6006030	180300	60000	3	114	127	38	6.41735e+02	6.41738e+02	2.0e-06	3.4e-07	2.1e-06	16.2	34.9
6002040	180300	20000	4	392	415	161	1.04527e+03	1.04527e+03	1.5e-07	1.0e-08	2.2e-07	42.9	93.5
7005030	245350	50000	3	197	205	7	3.13202e+02	3.13203e+02	2.1e-07	7.5e-09	1.1e-06	37.5	87.0
7007030	245350	70000	3	122	134	41	-3.69558e+02	-3.69558e+02	5.4e-07	1.4e-08	4.8e-07	21.5	47.6
7009030	245350	90000	3	110	136	37	-2.67562e+01	-2.67554e+01	1.6e-06	2.8e-07	1.5e-05	20.9	45.0
8007030	320400	70000	3	177	188	57	2.33140e+03	2.33140e+03	4.1e-07	1.2e-08	1.4e-07	36.7	85.3
80010030	320400	100000	3	116	129	9	2.25929e+03	2.25929e+03	1.3e-07	6.1e-09	2.9e-08	33.2	74.7
80011030	320400	110000	3	113	127	17	1.85792e+03	1.85792e+03	1.5e-07	3.2e-08	1.3e-07	31.0	68.5
90010030	405450	100000	3	150	161	16	9.54223e+02	9.54223e+02	2.9e-07	6.5e-09	2.9e-07	50.8	118.8
90014030	405450	140000	3	113	126	16	2.31983e+03	2.31983e+03	1.8e-07	2.5e-08	5.2e-08	40.9	96.7
100010030	500500	100000	3	204	213	56	3.09636e+03	3.09636e+03	2.2e-07	7.0e-09	4.2e-08	77.4	186.8
100015030	500500	150000	3	119	131	27	1.05289e+03	1.05289e+03	6.7e-07	2.0e-08	2.7e-07	50.3	117.5

Table 1: First-Algorithm-Performance with $t = 0.95$, $tol = 10^{-6}$ and $Z^{(0)} = 0$ on random sparse SDPs considered in [56]

Results where the boundary point method performed "better" than our first method are marked in red in the columns R_P and **time**. Half (12/24) of the considered problems were solved in less time (factor 1-2.5) and to higher accuracy by our method. Three problems were solved in more time but still to higher accuracy by our method. In these cases an exact Cholesky factor was relatively cheap (see [69]), explaining the time advantage of the boundary point method. The remaining 9 problems were solved less accurately, but also in less time by our method. A lack of primal infeasibility of the very first problem stands against a factor 3 in time. All eight remaining problems were solved within a range of $1.2 * 10^{-6} - 3.1 * 10^{-6}$ of primal infeasibility. Note that the relatively low number of CG iterations is more a result of our starting point than of our preconditioner.

seed	$(n^2+n)/2$	m	p	it	$C \bullet X$	$b^T y$	R_P	R_D	gap	time	cpu
3002030	45150	20000	3	163	7.61352e+02	7.61352e+02	9.4e-07	3.2e-08	3.0e-07	10.7	18.9
3002530	45150	25000	3	244	7.38384e+01	7.38385e+01	9.3e-07	4.9e-08	8.2e-07	26.9	43.2
3001040	45150	10000	4	151	1.65975e+02	1.65975e+02	9.6e-07	7.5e-08	1.3e-06	12.3	21.4
4003030	80200	30000	3	143	1.07214e+03	1.07214e+03	9.9e-07	2.8e-08	9.4e-07	12.3	23.6
4004030	80200	40000	3	193	8.05770e+02	8.05769e+02	9.4e-07	3.4e-08	7.8e-07	40.0	70.8
4001540	80200	15000	4	171	-6.55000e+02	-6.54998e+02	9.8e-07	1.2e-07	1.3e-06	22.1	40.5
5003030	125250	30000	3	153	1.10763e+03	1.10763e+03	9.5e-07	8.6e-08	4.1e-07	15.2	33.8
5004030	125250	40000	3	137	8.16610e+02	8.16611e+02	9.4e-07	3.6e-08	3.2e-07	16.9	35.3
5005030	125250	50000	3	149	3.64946e+02	3.64945e+02	1.0e-06	2.4e-08	7.4e-07	29.0	54.5
5002040	125250	20000	4	201	3.28005e+02	3.28011e+02	9.6e-07	2.2e-07	9.4e-06	35.4	66.1
6004030	180300	40000	3	154	3.06618e+02	3.06618e+02	9.7e-07	8.5e-08	5.4e-07	22.0	49.7
6005030	180300	50000	3	143	-3.86414e+02	-3.86413e+02	1.0e-06	6.0e-08	5.9e-07	20.8	46.5
6006030	180300	60000	3	138	6.41737e+02	6.41737e+02	9.7e-07	3.1e-08	2.5e-08	26.1	54.2
6002040	180300	20000	4	231	1.04527e+03	1.04528e+03	9.8e-07	4.0e-07	7.5e-06	37.6	78.1
7005030	245350	50000	3	168	3.13203e+02	3.13206e+02	9.9e-07	1.2e-07	5.2e-06	31.7	74.0
7007030	245350	70000	3	138	-3.69559e+02	-3.69559e+02	9.6e-07	4.3e-08	9.6e-09	31.5	70.0
7009030	245350	90000	3	142	-2.67572e+01	-2.67555e+01	9.2e-07	2.1e-08	3.1e-05	61.1	121.4
8007030	320400	70000	3	161	2.33140e+03	2.33140e+03	9.6e-07	8.5e-08	4.9e-07	40.1	94.2
80010030	320400	100000	3	135	2.25929e+03	2.25929e+03	9.9e-07	3.1e-08	2.0e-07	49.3	105.9
80011030	320400	110000	3	140	1.85792e+03	1.85792e+03	9.4e-07	2.3e-08	3.0e-07	66.4	134.9
90010030	405450	100000	3	149	9.54223e+02	9.54224e+02	9.9e-07	5.8e-08	3.8e-07	53.0	124.6
90014030	405450	140000	3	139	2.31983e+03	2.31983e+03	9.5e-07	2.4e-08	1.4e-07	107.3	230.8
100010030	500500	100000	3	172	3.09636e+03	3.09637e+03	9.9e-07	1.1e-07	5.9e-07	73.8	180.3
100015030	500500	150000	3	138	1.05288e+03	1.05289e+03	9.6e-07	3.4e-08	8.2e-07	85.2	193.4

Table 2: Boundary-Method-Performance with $\sigma = 0.1$ and $tol = 10^{-6}$ on random sparse SDPs considered in [56]

As mentioned before, for sake of completeness, we include the performance results of SDPNAL, again emphasizing that they are not representative in any average way.

seed	$(n^2+n)/2$	m	p	it	itsub	CGit	$C \bullet X$	$b^T y$	R_P	R_D	gap	time	cpu
3002030	45150	20000	3	7	46	4e+03	7.61352e+02	7.61371e+02	5.1e-07	6.6e-07	1.2e-05	31.0	85.7
3002530	45150	25000	3	5	61	5391	7.38394e+01	7.38586e+01	4.9e-07	6.8e-07	1.3e-04	51.3	156.6
3001040	45150	10000	4	11	51	985	1.65975e+02	1.66002e+02	2.0e-07	8.5e-07	8.3e-05	10.4	30.8
4003030	80200	30000	3	8	46	3219	1.07214e+03	1.07216e+03	8.4e-07	4.9e-07	9.2e-06	52.4	145.9
4004030	80200	40000	3	6	74	6206	8.06726e+02	8.05803e+02	8.2e-06	7.7e-07	5.7e-04	114.2	322.9
4001540	80200	15000	4	11	48	742	-6.55000e+02	-6.54964e+02	7.1e-07	8.0e-07	2.8e-05	15.8	45.8
5003030	125250	30000	3	10	49	2155	1.10763e+03	1.10766e+03	2.3e-07	7.1e-07	1.5e-05	47.1	154.6
5004030	125250	40000	3	9	45	2001	8.16612e+02	8.16655e+02	9.4e-07	8.1e-07	2.7e-05	50.6	147.3
5005030	125250	50000	3	7	45	3e+03	3.64946e+02	3.65005e+02	1.8e-07	9.5e-07	8.2e-05	90.3	267.3
5002040	125250	20000	4	11	52	819	3.28004e+02	3.28050e+02	3.2e-07	7.5e-07	7.0e-05	25.7	74.9
6004030	180300	40000	3	11	51	1859	3.06617e+02	3.06660e+02	1.9e-07	7.2e-07	7.0e-05	64.9	203.3
6005030	180300	50000	3	10	52	3e+03	-3.86413e+02	-3.86372e+02	2.5e-07	6.1e-07	5.3e-05	132.2	381.8
6006030	180300	60000	3	9	48	2374	6.41737e+02	6.41806e+02	1.1e-07	9.2e-07	5.4e-05	94.9	289.7
6002040	180300	20000	4	13	55	728	1.04527e+03	1.04530e+03	1.6e-07	5.5e-07	1.5e-05	29.6	86.3
7005030	245350	50000	3	11	51	1582	3.13203e+02	3.13260e+02	1.2e-07	7.7e-07	9.0e-05	85.8	265.3
7007030	245350	70000	3	10	49	2118	-3.69558e+02	-3.69504e+02	2.1e-07	6.3e-07	7.3e-05	127.0	365.1
7009030	245350	90000	3	9	46	2798	-2.67550e+01	-2.66947e+01	8.3e-07	6.4e-07	1.1e-03	202.0	615.5
8007030	320400	70000	3	11	51	1746	2.33140e+03	2.33146e+03	1.6e-07	7.2e-07	1.4e-05	135.2	408.2
80010030	320400	100000	3	10	52	3081	2.25929e+03	2.25935e+03	7.0e-07	5.2e-07	1.3e-05	259.4	811.3
80011030	320400	110000	3	10	51	2730	1.85792e+03	1.85796e+03	1.5e-07	3.8e-07	1.1e-05	260.3	787.3
90010030	405450	100000	3	11	52	2008	9.54223e+02	9.54298e+02	1.3e-07	6.3e-07	3.9e-05	223.5	692.6
90014030	405450	140000	3	9	49	2782	2.31983e+03	2.31997e+03	2.4e-07	9.2e-07	2.9e-05	356.8	1117.4
100010030	500500	100000	3	11	51	2e+03	3.09636e+03	3.09645e+03	6.9e-07	7.1e-07	1.5e-05	250.1	806.5
100015030	500500	150000	3	10	50	2956	1.05289e+03	1.05302e+03	1.7e-07	8.4e-07	6.3e-05	455.4	1441.7

Table 3: SDPNAL-Performance with $tol = 10^{-6}$ on random sparse SDPs considered in [56]

For our second test the tolerance was dramatically reduced to $tol = 10^{-11}$, for which the boundary point method encountered issues, probably due to roundoff errors. Therefore we only include the performance results of the boundary point method for $tol = 10^{-10}$.

seed	$(n^2+n)/2$	m	p	it	CGit	up	$C \bullet X$	$b^T y$	R_P	R_D	gap	time	cpu
3002030	45150	20000	3	239	405	182	7.61352e+02	7.61352e+02	5.0e-11	9.1e-14	4.9e-11	8.6	15.3
3002530	45150	25000	3	809	1374	643	7.38384e+01	7.38384e+01	2.6e-11	2.9e-14	4.8e-11	37.6	56.5
3001040	45150	10000	4	404	652	299	1.65974e+02	1.65974e+02	2.2e-11	7.1e-13	4.8e-11	16.9	26.1
4003030	80200	30000	3	217	245	147	1.07214e+03	1.07214e+03	2.9e-11	1.0e-11	2.0e-11	10.3	20.1
4004030	80200	40000	3	419	712	306	8.05769e+02	8.05769e+02	3.5e-11	3.0e-14	2.2e-11	29.3	48.9
4001540	80200	15000	4	528	812	357	-6.55000e+02	-6.55000e+02	2.5e-12	9.8e-14	1.0e-12	29.2	51.2
5003030	125250	30000	3	390	401	286	1.10763e+03	1.10763e+03	2.5e-11	8.2e-13	4.3e-12	23.3	48.6
5004030	125250	40000	3	242	255	161	8.16611e+02	8.16611e+02	4.7e-12	8.4e-14	2.8e-12	16.9	34.9
5005030	125250	50000	3	209	242	131	3.64945e+02	3.64945e+02	1.2e-11	1.6e-12	2.0e-11	16.4	33.1
5002040	125250	20000	4	700	1044	535	3.28004e+02	3.28004e+02	3.1e-12	1.5e-13	1.5e-12	45.0	85.1
6004030	180300	40000	3	429	440	332	3.06617e+02	3.06617e+02	1.2e-11	4.2e-13	3.2e-12	33.1	70.6
6005030	180300	50000	3	323	335	233	-3.86414e+02	-3.86414e+02	2.9e-12	6.7e-14	4.1e-12	30.5	64.3
6006030	180300	60000	3	226	239	148	6.41737e+02	6.41737e+02	1.8e-10	7.6e-12	1.6e-10	21.9	47.2
6002040	180300	20000	4	1023	1046	780	1.04527e+03	1.04527e+03	2.0e-12	1.4e-13	2.2e-12	75.0	158.0
7005030	245350	50000	3	497	505	291	3.13203e+02	3.13203e+02	5.1e-12	1.9e-13	1.8e-11	63.9	144.0
7007030	245350	70000	3	291	303	207	-3.69559e+02	-3.69559e+02	1.7e-11	3.2e-13	8.7e-12	36.4	78.7
7009030	245350	90000	3	219	237	144	-2.67555e+01	-2.67555e+01	1.5e-11	5.5e-12	1.3e-10	32.1	67.6
8007030	320400	70000	3	444	455	319	2.33140e+03	2.33140e+03	1.4e-11	4.3e-13	8.3e-13	68.7	154.3
80010030	320400	100000	3	250	263	140	2.25929e+03	2.25929e+03	1.2e-11	1.8e-13	3.8e-12	50.6	113.2
80011030	320400	110000	3	223	232	125	1.85792e+03	1.85792e+03	6.5e-12	6.8e-13	2.9e-12	45.6	100.2
90010030	405450	100000	3	371	382	229	9.54223e+02	9.54223e+02	1.1e-11	2.6e-13	9.8e-12	87.4	199.4
90014030	405450	140000	3	223	233	124	2.31983e+03	2.31983e+03	9.8e-12	5.4e-13	5.6e-12	62.1	141.3
100010030	500500	100000	3	515	524	361	3.09636e+03	3.09636e+03	5.2e-12	1.7e-13	3.0e-13	131.7	308.9
100015030	500500	150000	3	274	286	178	1.05289e+03	1.05289e+03	3.6e-12	5.2e-14	5.9e-13	84.5	190.6

Table 4: First-Algorithm-Performance with $t = 0.95$, $tol = 10^{-11}$ and $Z^{(0)} = 0$ on random sparse SDPs considered in [56]

seed	$(n^2+n)/2$	**m**	**p**	**it**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
3002030	45150	20000	3	610	7.61352e+02	7.61352e+02	8.1e-11	1.0e-10	1.7e-10	33.7	58.6
3002530	45150	25000	3	551	7.38384e+01	7.38384e+01	9.5e-11	9.9e-11	3.8e-10	56.5	83.8
3001040	45150	10000	4	1554	1.65974e+02	1.65974e+02	8.4e-11	1.0e-10	2.5e-09	103.0	165.2
4003030	80200	30000	3	733	1.07214e+03	1.07214e+03	9.2e-11	9.9e-11	8.9e-11	58.4	110.3
4004030	80200	40000	3	542	8.05769e+02	8.05769e+02	8.8e-11	9.9e-11	6.3e-11	103.9	156.8
4001540	80200	15000	4	2001	-6.55000e+02	-6.55000e+02	1.1e-10	1.3e-10	8.3e-10	227.7	370.9
5003030	125250	30000	3	1462	1.10763e+03	1.10763e+03	9.5e-11	1.0e-10	4.0e-10	132.1	298.3
5004030	125250	40000	3	993	8.16611e+02	8.16611e+02	8.2e-11	1.0e-10	5.9e-11	105.1	221.8
5005030	125250	50000	3	673	3.64945e+02	3.64945e+02	9.6e-11	1.0e-10	5.8e-10	115.3	204.4
5002040	125250	20000	4	2001	3.28004e+02	3.28004e+02	7.4e-10	8.8e-10	2.6e-08	321.2	555.8
6004030	180300	40000	3	1656	3.06617e+02	3.06617e+02	8.3e-11	1.0e-10	1.7e-09	215.2	494.5
6005030	180300	50000	3	1333	-3.86414e+02	-3.86414e+02	9.1e-11	1.0e-10	6.4e-10	183.2	409.5
6006030	180300	60000	3	915	6.41737e+02	6.41737e+02	9.5e-11	1.0e-10	6.8e-11	153.5	312.1
6002040	180300	20000	4	2001	1.04527e+03	1.04527e+03	3.9e-09	4.5e-09	5.2e-08	313.5	646.5
7005030	245350	50000	3	1986	3.13203e+02	3.13203e+02	8.2e-11	1.0e-10	3.0e-09	355.7	830.9
7007030	245350	70000	3	1213	-3.69559e+02	-3.69559e+02	8.6e-11	1.0e-10	8.6e-11	242.6	536.2
7009030	245350	90000	3	778	-2.67555e+01	-2.67555e+01	9.4e-11	1.0e-10	2.5e-09	271.1	481.0
8007030	320400	70000	3	1765	2.33140e+03	2.33140e+03	9.1e-11	1.0e-10	4.0e-10	428.6	1002.3
80010030	320400	100000	3	1052	2.25929e+03	2.25929e+03	8.3e-11	1.0e-10	3.0e-11	340.0	702.3
80011030	320400	110000	3	874	1.85792e+03	1.85792e+03	9.4e-11	1.0e-10	1.5e-10	373.1	697.7
90010030	405450	100000	3	1591	9.54223e+02	9.54223e+02	8.5e-11	1.0e-10	7.9e-10	529.4	1234.1
90014030	405450	140000	3	927	2.31983e+03	2.31983e+03	9.6e-11	1.0e-10	1.0e-10	550.7	1018.2
100010030	500500	100000	3	2001	3.09636e+03	3.09636e+03	1.0e-10	1.1e-10	5.3e-10	833.7	2017.6
100015030	500500	150000	3	1214	1.05289e+03	1.05289e+03	8.5e-11	1.0e-10	3.0e-10	640.3	1374.9

Table 5: Boundary-Method-Performance with $\sigma = 0.1$, $tol = 10^{-10}$ and $maxit = 2000$ on random sparse SDPs considered in [56]

It is not intended that these very limited results are used for ranking the methods. Although this first implementation seems to be one step ahead for these specific random sparse SDPs, there is no pretense that this would be the case in general. Some additional testing especially on SDP-relaxations coming from Max-Clique/Max-Stable-Set problems, suggested that the method is often, although not always, competitive due to its relatively low complexity per iteration and especially due to the update strategy (112) for $|Z|_{\mathcal{K}}$ described above. I do not think that this simple update strategy has been exploited for SDPs in any implementation before, therefore it might help researchers tuning their implementations in the future. There are several observations to be made while working with such a code that are not entirely obvious from the tables above. First of all the residuals $\left\|\nabla f(Z^{(k)})\right\|$ decrease monotonically with k, unless the approximation $\Delta Z^{(k)} \approx -2\mathcal{H}[\nabla f(Z^{(k)})]$ is "too inexact", independent of the step length $t \in [0,1]$. The second observation that one can make is that choosing step lengths smaller than $\frac{1}{2}$ often leads to much much slower convergence than step lengths larger or equal to $\frac{1}{2}$. Especially values close to, but strictly smaller than one work surprisingly well in practice (we will come back to this phenomenon later). Point three is not really an observation that can be made from looking at our implementation: If $Z^{(*)} \in E$ satisfies $\nabla f(Z^{(*)}) = 0$ then the distance $\left\|Z^{(k)} - Z^{(*)}\right\|$ will decrease monotonically in k, if the approximation $\Delta Z^{(k)} \approx -2\mathcal{H}[\nabla f(Z^{(k)})]$ is sufficiently exact. We especially acknowledge the numerical stability of inexact versions of (109). Unfortunately a proper mathematical treatment of inexact methods can (and usually does) get complicated very quickly. Our goal in this subsection is to give a simple mathematical introduction, merely an impression about what is to come. We will therefore prove all results only for an exact iteration (i.e. $\Delta Z^{(k)} = -2\mathcal{H}[\nabla f(Z^{(k)})]$) making all considerations much simpler. Interesting and extensive analysis of the inexact iteration can be found in [52] and the recent paper [5]. Now let us give our observations a mathematical foundation and state and prove the following lemma (with essentially well known results) concerning the step lengths $t_k \in \mathbb{R}$:

Lemma 4.1. *For all $Z \in E$ the direction $\Delta Z = -2\mathcal{H}[\nabla f(Z)]$ is a (weak) descent direction* *of the gradient norms, specifically for all $t \in [0,1]$ we obtain*

$$\|\nabla f(Z + t\Delta Z)\| \leq \|\nabla f(Z)\|.$$

Furthermore, if $Z^{(*)} \in E$ satisfies $\nabla f(Z^{(*)}) = 0$, then

$$\left\|Z + t\Delta Z - Z^{(*)}\right\|^2 \leq \left\|Z - Z^{(*)}\right\|^2 - 4t(1-t)\underbrace{\langle \mathcal{H}[\nabla f(Z)], Z - Z^{(*)}\rangle}_{\geq 0}$$

For $Z \in E$ with $\nabla f(Z) \neq 0$ the step length minimizing the distance to $Z^{(*)} \in E$ is lower bounded by $\frac{1}{2}$, i.e.

$$\underset{t\in\mathbb{R}}{\mathit{argmin}} \left\|Z + t\Delta Z - Z^{(*)}\right\| = \frac{1}{2}\underbrace{\tfrac{\langle \nabla f(Z), \mathcal{H}[Z - Z^{(*)}]\rangle}{\|\nabla f(Z)\|^2}}_{\geq 1} \geq \frac{1}{2} \tag{114}$$

Proof. By equation (22) we have the equality

$$\nabla f(Z - 2t\mathcal{H}[\nabla f(Z)]) = (1-t)\nabla f(Z) - \tfrac{1}{2}(|Z - 2t\mathcal{H}[\nabla f(Z)]|_{\mathcal{K}} - |Z|_{\mathcal{K}})$$

for all $t \in \mathbb{R}$. For $t \in [0,1]$ this together with the triangle inequality and Lipschitz-Continuity implies

$$\begin{aligned}\|\nabla f(Z - 2t\mathcal{H}[\nabla f(Z)])\| &\leq (1-t)\|\nabla f(Z)\| + \tfrac{1}{2}\||Z - 2t\mathcal{H}[\nabla f(Z)]|_{\mathcal{K}} - |Z|_{\mathcal{K}}\| \\ &\leq (1-t)\|\nabla f(Z)\| + \tfrac{1}{2}\|2t\mathcal{H}[\nabla f(Z)]\| = \|\nabla f(Z)\|\end{aligned}$$

i.e. monotone decrease. Note that there is a different way for $t \in (0,1]$ to prove the above , i.e.

$$\begin{aligned}\|\nabla f(Z - 2t\mathcal{H}[\nabla f(Z)])\|^2 &= \|\nabla f(Z)\|^2 - \tfrac{(1-t)}{t}\langle \nabla f(Z - 2t\mathcal{H}[\nabla f(Z)]) - \nabla f(Z), -2t\nabla f(Z)\rangle \\ &\quad + \tfrac{1}{4}\||Z - 2t\mathcal{H}[\nabla f(Z)]|_{\mathcal{K}} - |Z|_{\mathcal{K}}\|^2 - \tfrac{1}{4}\|2t\nabla f(Z)\|^2 \\ &\leq \|\nabla f(Z)\|^2 - \frac{(1-t)}{t}\|\nabla f(Z - 2t\mathcal{H}[\nabla f(Z)]) - \nabla f(Z)\|^2\end{aligned} \tag{115}$$

which is probably more common knowledge as it often appears similarly when proving convergence of gradients method for convex functions. Let now $Z^{(*)} \in E$ satisfy $\nabla f(Z^{(*)}) = 0$ then $\forall Z \in E$

$$\begin{aligned}&\left\|Z - 2t\mathcal{H}[\nabla f(Z)] - Z^{(*)}\right\|^2 \\ &= \left\|Z - Z^{(*)}\right\|^2 + 4t^2 \underbrace{\|\nabla f(Z)\|^2}_{\leq \langle \nabla f(Z), \mathcal{H}[Z - Z^{(*)}]\rangle} - 4t\langle \nabla f(Z), \mathcal{H}[Z - Z^{(*)}]\rangle \\ &\leq \left\|Z - Z^{(*)}\right\|^2 - 4t(1-t)\langle \nabla f(Z), \mathcal{H}[Z - Z^{(*)}]\rangle\end{aligned} \tag{116}$$

is satisfied. Another way to express this monotone decrease comes from the equation

$$\begin{aligned}&\left\|Z - 2t\mathcal{H}[\nabla f(Z)] - Z^{(*)}\right\|^2 \\ &= \left\|Z - Z^{(*)}\right\|^2 - \underbrace{4t(1-t)\|\nabla f(Z)\|^2}_{\geq 0} + \underbrace{4t(\|\nabla f(Z)\|^2 - \langle \nabla f(Z), \mathcal{H}[Z - Z^{(*)}]\rangle)}_{\leq 0}\end{aligned} \tag{117}$$

which we can also use to easily derive the optimal step length (i.e. $t^* \in \mathbb{R}$ minimizing the right hand side) for given $Z \in E : \nabla f(Z) \neq 0$ by setting the derivative in t of the right hand side equal to zero

$$8t^* \|\nabla f(Z)\|^2 - 4\langle \nabla f(Z), \mathcal{H}[Z - Z^{(*)}]\rangle \stackrel{!}{=} 0 \tag{118}$$

we obtain

$$t^* := \frac{1}{2} \underbrace{\frac{\langle \nabla f(Z), \mathcal{H}[Z - Z^{(*)}]\rangle}{\|\nabla f(Z)\|^2}}_{\geq 1} \geq \frac{1}{2} \tag{119}$$

which also minimizes the distance $\left\|Z + t\Delta Z - Z^{(*)}\right\|$ in t. ■

Above lemma readily explains some of the observations that we made before and it suffices to establish the following convergence result (which is not new, a proof can for example be found (with different notation) in [14]):

Proposition 4.2. *If there exists $Z^{(*)} \in E$ satisfying $\nabla f(Z^{(*)}) = 0$ and the step lengths satisfy $t_k \in (0,1)$ as well as $\lim_{k\to\infty} \sum_{i=0}^{k} 4t_i(1 - t_i) = \infty$. Then for any $Z^{(0)} \in E$ the sequence defined by*

$$Z^{(k+1)} = Z^{(k)} - 2t_k\mathcal{H}[\nabla f(Z^{(k)})] \;\forall k \in \mathbb{N} \tag{120}$$

will converge to some $Z^{**} \in E$ satisfying $\nabla f(Z^{**}) = 0$.

Proof. By the previous lemma we have for every $k \in \mathbb{N}$

$$\begin{aligned}
&\sum_{i=0}^{k} 4t_i(1 - t_i) \left\|\nabla f(Z^{(k)})\right\|^2 \leq \sum_{i=0}^{k} 4t_i(1 - t_i) \left\|\nabla f(Z^{(i)})\right\|^2 \\
\leq &\sum_{i=0}^{k} 4t_i(1 - t_i)\langle \mathcal{H}[\nabla f(Z^{(i)})], Z^{(i)} - Z^{(*)}\rangle \leq \sum_{i=0}^{k}(\left\|Z^{(i)} - Z^{(*)}\right\|^2 - \left\|Z^{i+1} - Z^{(*)}\right\|^2) \\
=&(\left\|Z^{(0)} - Z^{(*)}\right\|^2 - \left\|Z^{(k+1)} - Z^{(*)}\right\|^2) \leq \left\|Z^{(0)} - Z^{(*)}\right\|^2
\end{aligned} \tag{121}$$

implying

$$\left\|\nabla f(Z^{(k)})\right\|^2 \leq \tfrac{\left\|Z^{(0)} - Z^{(*)}\right\|^2}{\sum_{i=0}^{k} 4t_i(1-t_i)} \tag{122}$$

and therefore $\lim_{k\to\infty} \left\|\nabla f(Z^{(k)})\right\| = 0$. Since the sequence $\{Z^{(k)}\}_{k\in\mathbb{N}} \subset E$ is bounded and E is a finite dimensional Euclidean space this implies its convergence to some $Z^{**} \in E$ satisfying $\nabla f(Z^{**}) = 0$. ■

Note that above proof comes with a slightly better error bound for the residuals: Let us denote the distance function with $dist(Z, \mathcal{N}) := ||Z - \Pi_{\mathcal{N}}(Z)||$ where $\Pi_{\mathcal{N}}(Z) \in E$ denotes the orthogonal projections onto the set of zeros $\mathcal{N} := \{Z \in E \mid \nabla f(Z) = 0\}$. Then

$$\left\|\nabla f(Z^{(k)})\right\|^2 \leq \tfrac{dist(Z^{(0)}, \mathcal{N})^2}{\sum_{i=0}^{k} 4t_i(1-t_i)} \tag{123}$$

holds true. If for example there exists $\frac{1}{2} \geq \rho > 0$ such that the step lengths $t_k \in [\rho, 1-\rho] \; \forall k \in \mathbb{N}_0$ are bounded away from zero and one, then above bound implies asymptotically $\left\|\nabla f(Z^{(k)}\right\|^2 \in O(k^{-1})$ for $k \to \infty$ whenever $\mathcal{N} \neq \emptyset$. We can improve this bound in finite dimension from $O(k^{-1})$ to $o(k^{-1})$ for $k \to \infty$ as is shown in the following corollary:

Corollary 4.3. *If there exists $Z^{(*)} \in E$ satisfying $\nabla f(Z^{(*)}) = 0$ and there exists $\frac{1}{2} \geq \rho > 0$ such that the step lengths satisfy $t_k \in [\rho, 1-\rho] \; \forall k \in \mathbb{N}_0$. Then for any $Z^{(0)} \in E$ the sequence defined by*

$$Z^{(k+1)} = Z^{(k)} - 2t_k \mathcal{H}[\nabla f(Z^{(k)})] \; \forall k \in \mathbb{N} \tag{124}$$

will satisfy

$$dist(Z^{(k+1)}, \mathcal{N})^2 \leq dist(Z^{(k)}, \mathcal{N})^2 - 4t_k(1-t_k) \left\|\nabla f(Z^{(k)})\right\|^2 \tag{125}$$

and

$$\left\|\nabla f(Z^{(k)})\right\|^2 \in o(k^{-1}) \quad \text{for} \quad k \to \infty \tag{126}$$

Proof. The first part of our claim follows from the definition of the distance function and our previous lemma

$$\begin{aligned}
dist(Z^{(k+1)}, \mathcal{N})^2 = &\left\|Z^{(k+1)} - \Pi_{\mathcal{N}}(Z^{(k+1)})\right\|^2 \leq \left\|Z^{(k+1)} - \Pi_{\mathcal{N}}(Z^{(k)})\right\|^2 \\
\leq &\left\|Z^{(k)} - \Pi_{\mathcal{N}}(Z^{(k)})\right\|^2 - 4t_k(1-t_k)\langle \mathcal{H}[\nabla f(Z^{(k)}]), Z^{(k)} - \Pi_{\mathcal{N}}(Z^{(k)})\rangle \\
= &dist(Z^{(k)}, \mathcal{N}) - 4t_k(1-t_k)\langle \mathcal{H}[\nabla f(Z^{(k)})], Z^{(k)} - \Pi_{\mathcal{N}}(Z^{(k)})\rangle \\
\leq &dist(Z^{(k)}, \mathcal{N}) - 4t_k(1-t_k) \left\|\nabla f(Z^{(k)})\right\|^2.
\end{aligned} \tag{127}$$

This implies for $k \geq 1$

$$2k \left\|\nabla f(Z^{2k})\right\|^2 \frac{\sum_{i=k}^{2k} 2t_i(1-t_i)}{k} \leq \sum_{i=k}^{2k} 4t_i(1-t_i) \left\|\nabla f(Z^{(i)})\right\|^2 \leq dist(Z^{2k+1}, \mathcal{N})^2 - dist(Z^{(k)}, \mathcal{N})^2 \overset{k\to\infty}{\to} 0 \tag{128}$$

and therefore proves our second claim, since our boundness assumption on t_k translates to the existence of two positive constants $0 < \rho_1 \leq \rho_2$ such that $\frac{\sum_{i=k}^{2k} 2t_i(1-t_i)}{k} \in [\rho_1, \rho_2] \; \forall k \in \mathbb{N}$ holds true. ■

There are two problems regarding the above error bounds: First note that the term $\frac{1}{\sum_{i=0}^{k} 4t_i(1-t_i)}$ can very well be strictly larger than one, but the norm of the residuals of (109) are always monotone in k implying that $\left\|\nabla f(Z^{(k)})\right\| \leq dist(Z^{(0)}, \mathcal{N})$ holds true $\forall k \in \mathbb{N}_0$. In other words, bound (123) may not hold any information. The second problem is the gap between our bounds and our numerical experiments: To obtain the results of Table (1) we chose $t_k = t = 0.95$ and observed a maximal number of iterations $k_{max} := 392$, i.e. $\frac{1}{\sum_{i=1}^{k_{max}} 4t_i(1-t_i)} = \frac{1}{0.19(k_{max}+1)} = \frac{100}{19(392+1)} = \frac{100}{7467} \approx 0.0134$. This means $\left\|\nabla f(x_{k_{max}}))\right\| \leq \sqrt{\frac{100}{7467}} dist(Z^{(0)}, \mathcal{N}) = \underbrace{\sqrt{\frac{100}{7467}}}_{\approx 0.115725} dist(0, \mathcal{N})$. Note that our normalization implies $\left\|Z^{(*)}\right\| \geq 1 \ \forall Z^{(*)} \in \mathcal{N}$, i.e. $dist(Z^{(0)}, \mathcal{N}) \geq 1$. Our theoretical bound is therefore larger than $\sqrt{\frac{100}{7467}} \approx 0.115725$ which is several orders of magnitude larger than the observed value of $\|\nabla f(x_{k_{max}})\| \approx 10^{-6}$. Obviously giving one set of example is not representative, since our problems could simply be "easy" to solve. Indeed if one replaces $\mathcal{H}\nabla f$ by some firmly nonexpansive operator F, then it is known that for constant step length the (adapted) bound (123) can not be improved by more than a constant (see e.g. [1]). However although we encountered some "harder" examples in our numerical tests, there always seemed to be a gap between the theoretical bound and the observed numerical behavior. As mentioned earlier, this gap initially led to the question of a possibly overlooked inner structure of $\mathcal{H}[\nabla f(Z)]$. We now consider two types of examples, to show that this was not the case. The reason for stating them as theorems will unveil itself in corollary 4.7 and in the next subsection, where we will see that they represent certain "worst-case" examples.

Theorem 4.4 (Tightness of complexity for constant step length part 1)**.** *Let $t \in [\frac{1}{2}, 1)$ be fixed. Then for every $\kappa \in \mathbb{N}$ satisfying $\kappa + 1 > \frac{1}{4t(1-t)}$ there exists a problem (P_κ) taking the form (1) for a two dimensional Euclidean space E such that the via (18) associated function $\nabla f_\kappa : E \to E$ has a unique zero $Z^{(*)} \in E$ and the iterates defined via $Z^{(k+1)} := Z^{(k)} - 2t\mathcal{H}_\kappa[\nabla f_\kappa(Z^{(k)})]$ will satisfy*

$$\left\|\nabla f_\kappa(Z^{(k)})\right\|^2 = \frac{1}{\kappa+1}(\frac{\kappa}{\kappa+1})^k \frac{\left\|Z^{(0)} - Z^{(*)}\right\|^2}{4t(1-t)} \qquad \forall k \in \mathbb{N}_0 \tag{129}$$

for every $Z^{(0)} \in E$. Specifically

$$\left\|\nabla f_\kappa(Z^{(\kappa)})\right\|^2 = \frac{1}{\kappa+1}(\frac{\kappa}{\kappa+1})^\kappa \frac{\left\|Z^{(0)} - Z^{(*)}\right\|^2}{4t(1-t)} > \frac{1}{\kappa+1}\frac{1}{\exp(1)}\frac{\left\|Z^{(0)} - Z^{(*)}\right\|^2}{4t(1-t)} \tag{130}$$

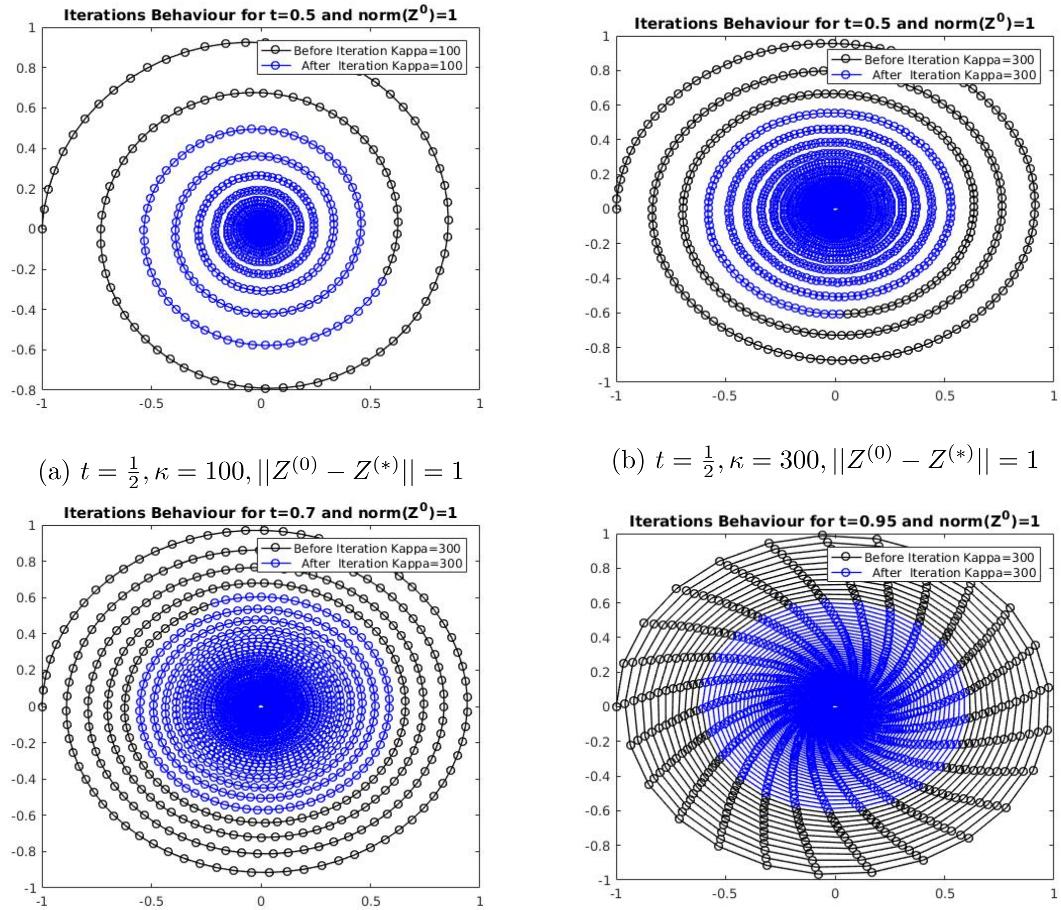
Remark 4.5. *Now, one might look at the result above and be calmed by the term $(\frac{\kappa}{\kappa+1})^k$ which, afterall, will for $k \to \infty$ converge to zero quite rapidly. This is however slightly misleading: Yes, after the κth iterate the convergence turns out to be linear, but until then it is sublinear. In other words the "local" regime of quick convergence does exist (in this example), but we have to endure the "global" convergence regime first: not infinitely, but arbitrarily long depending on κ. In fact note that $(\frac{\kappa}{\kappa+1})^\kappa \to \exp(-1) \approx 0.3679$. Note that, since our example is two dimensional, we can actually plot the iteration behavior, see figure 2.*

Proof. Let $t \in [\frac{1}{2}, 1)$ be fixed and $\kappa \in \mathbb{N}$ such that $\kappa + 1 > \frac{1}{4t(1-t)}$. We define

$$c_\kappa := 1 - \frac{1}{2(\kappa+1)t(1-t)} \in (-1, 1) \text{ and } \quad s_\kappa := \sqrt{1 - c_\kappa^2} \in (0, 1) \tag{131}$$

as well as the 2-dimensional vector $a_\kappa^T := [-\frac{1+c_\kappa}{s_\kappa}, 1]$. Let us consider the following optimization problem

$$\begin{aligned} \underset{X \in \mathbb{R}^2}{\text{minimize}} \quad & 0^T X \\ \text{subject to} \quad & a_\kappa^T X = 0 \\ & X \in \{0\} \times \mathbb{R} \end{aligned} \tag{132}$$



(a) $t = \frac{1}{2}, \kappa = 100, ||Z^{(0)} - Z^{(*)}|| = 1$

(b) $t = \frac{1}{2}, \kappa = 300, ||Z^{(0)} - Z^{(*)}|| = 1$

(c) $t = 0.7, \kappa = 300, ||Z^{(0)} - Z^{(*)}|| = 1$

(d) $t = 0.95, \kappa = 300, ||Z^{(0)} - Z^{(*)}|| = 1$

Figure 2: 2-dimensional examples with slow convergence (129) for different values of κ and t. The iterates (marked by small circles) rotate around the unique solution which is located in zero. The first κ iterates (black) converge sublinear. After κ iterates, the (faster) local convergence regime takes over and we observe linear convergence of the latter iterates (blue circles).

which clearly is of the form (1) for the (closed, convex) cone $\mathcal{K} = \{0\} \times \mathbb{R}$. Its unique optimal solution is trivially given by $X^{(*)} = 0 \in \mathbb{R}^2$. Now the dual problem takes the Form

$$\begin{aligned} \underset{Y \in \mathbb{R}^2,\ y \in \mathbb{R}}{\text{maximize}} \quad & 0y \\ \text{subject to} \quad & a_\kappa y - Y = 0 \\ & Y \in \mathcal{K}^P = \mathbb{R} \times \{0\} \end{aligned} \tag{133}$$

Note that its unique optimal solution is attained at $y^{(*)} = 0 \in \mathbb{R},\ Y^{(*)} = 0$. Here the generalized absolute value of some vector $Z = \begin{pmatrix} Z_1 & Z_2 \end{pmatrix}^T \in \mathbb{R}^2$ takes the form $|Z|_{\mathcal{K}} =$

$\begin{pmatrix} -Z_1 \\ Z_2 \end{pmatrix}$. Using the definition from equation (18) we obtain the function $\nabla f_\kappa : \mathbb{R}^2 \to \mathbb{R}^2$ via $\nabla f_\kappa(Z) := \frac{1}{2}(\mathcal{H}_\kappa Z - \begin{pmatrix} -Z_1 \\ Z_2 \end{pmatrix})$ for the orthogonal and symmetric matrix

$$\begin{aligned}\mathcal{H}_\kappa :=& I_2 - 2\frac{a_\kappa a_\kappa^T}{||a_\kappa||^2} = I - 2\frac{1}{1+(\frac{1+c_\kappa}{s_\kappa})^2}\begin{pmatrix} (\frac{1+c_\kappa}{s_\kappa})^2 & -\frac{1+c_\kappa}{s_\kappa} \\ -\frac{1+c_\kappa}{s_\kappa} & 1 \end{pmatrix} \\ =& I - \frac{1}{\frac{(1+c_\kappa)}{s_\kappa^2}}\begin{pmatrix} (\frac{1+c_\kappa}{s_\kappa})^2 & -\frac{1+c_\kappa}{s_\kappa} \\ -\frac{1+c_\kappa}{s_\kappa} & 1 \end{pmatrix} = \begin{pmatrix} -c_\kappa & s_\kappa \\ s_\kappa & c_\kappa \end{pmatrix}.\end{aligned}$$

Note that by construction $Z^{(*)} = 0 \in \mathbb{R}^2$ is the unique point that satisfies $\nabla f_\kappa(Z^{(*)}) = 0$. If we define the matrix $Q_\kappa := \mathcal{H}_\kappa \operatorname{Diag}([-1, 1]) = \begin{pmatrix} c_\kappa & s_\kappa \\ -s_\kappa & c_\kappa \end{pmatrix}$ then it is orthogonal, i.e. $Q_\kappa^T Q_\kappa = I$ and satisfies

$$Q_\kappa + Q_\kappa^T = 2c_\kappa I$$

and

$$(I-t(I-Q_\kappa^T))(I-t(I-Q_\kappa)) = (1-t)I+t^2Q_\kappa^TQ_\kappa+t(1-t)(Q_\kappa+Q_\kappa^T) = I(1-2t+2t^2+c_\kappa t(1-t)).$$

Furthermore we can write

$$\begin{aligned}\mathcal{H}_\kappa \nabla f_\kappa(Z) =& \frac{1}{2}(\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} - \begin{pmatrix} c_\kappa & s_\kappa \\ -s_\kappa & c_\kappa \end{pmatrix}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}) \\ =& \frac{1}{2}\begin{pmatrix} 1-c_\kappa & -s_\kappa \\ s_\kappa & 1-c_\kappa \end{pmatrix}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \frac{1}{2}(I - Q_\kappa)Z\end{aligned}$$

and for arbitrary $Z^{(0)} \in \mathbb{R}^2$ and $k \in \mathbb{N}$

$$\begin{aligned}Z^{(k)} =& Z^{(k-1)} - 2t\mathcal{H}_\kappa \nabla f_\kappa(Z^{(k-1)}) = (\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - t\begin{pmatrix} 1-c_\kappa & -s_\kappa \\ s_\kappa & 1-c_\kappa \end{pmatrix})Z^{(k-1)} \\ =& (\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - t\begin{pmatrix} 1-c_\kappa & -s_\kappa \\ s_\kappa & 1-c_\kappa \end{pmatrix})^k Z^{(0)} = (I - t(I-Q_\kappa))Z^{(0)}.\end{aligned}$$

By combining these equations we obtain

$$\begin{aligned}&||\nabla f_\kappa(Z^{(k)})||_2^2 = ||\tfrac{1}{2}(I-Q_\kappa)Z^{(k)}||_2^2 = \tfrac{1}{4}(Z^{(k)})^T(I - Q_\kappa^T - Q_\kappa + Q_\kappa^TQ_\kappa)Z^{(k)} \\ =& \tfrac{1}{4}(Z^{(k)})^T(I(2-2c_\kappa))Z^{(k)} = \frac{1-c_\kappa}{2}||Z^{(k)}||_2^2 = \frac{1-c_\kappa}{2}||(I-t(I-Q_\kappa))^kZ^{(0)}||_2^2 \\ =& \frac{1-c_\kappa}{2}(1-2t+2t^2+2c_\kappa t(1-t))^k||Z^{(0)}||^2 \\ =& \frac{1}{(\kappa+1)4t(1-t)}(1-2t+t^2+2t(1-t) - \frac{1}{\kappa+1})^k||Z^{(0)}||_2^2 \\ =& \frac{1}{(\kappa+1)4t(1-t)}(1-\frac{1}{\kappa+1})^k||Z^{(0)}||_2^2 = \frac{1}{\kappa+1}(\frac{\kappa}{\kappa+1})^k\frac{||Z^{(0)}||_2^2}{4t(1-t)} \\ =& \frac{1}{\kappa+1}(\frac{\kappa}{\kappa+1})^k\frac{||Z^{(0)} - Z^{(*)}||_2^2}{4t(1-t)}\end{aligned} \tag{134}$$

which yields the desired result. ■

Theorem 4.4 excludes the case of a step length t close to one for a small number of iterations. Indeed we made the assumption $\kappa+1 > \frac{1}{4t(1-t)}$, leaving the case $\kappa+1 \leq \frac{1}{4t(1-t)}$ open. Let us close this gap by giving the second example below:

Theorem 4.6 (Tightness of complexity for fixed step length part 2)**.** *Let $t \in (\frac{1}{2}, 1)$ be fixed. Then for every $\kappa \in \mathbb{N}$ satisfying $\kappa+1 \leq \frac{1}{4t(1-t)}$ there exists a problem (P) taking the Form (1) for a one dimensional Euclidean space E such that the via (18) associated function $\nabla f : E \to E$ has a unique zero $Z^{(*)} \in E$ and the iteration $Z^{(k+1)} = Z^{(k)} - 2t\mathcal{H}[\nabla f(Z^{(k)})]$ will satisfy*

$$\left\|\nabla f(Z^{(k)})\right\|^2 = (2t-1)^{2k} \left\|Z^{(0)} - Z^{(*)}\right\|^2 \qquad \forall k \in \mathbb{N}_0 \tag{135}$$

for every $Z^{(0)} \in E$. Specifically

$$\left\|\nabla f(Z^{(\kappa)})\right\|^2 = (2t-1)^{2\kappa} \left\|Z^{(0)} - Z^{(*)}\right\|^2 \geq (\frac{\kappa}{\kappa+1})^{\kappa} \left\|Z^{(0)} - Z^{(*)}\right\|^2 > \exp(-1) \left\|Z^{(0)} - Z^{(*)}\right\|^2 \tag{136}$$

Proof. Consider

$$\begin{aligned} \underset{X\in\mathbb{R}}{\text{minimize}} \quad & 0X \\ \text{subject to} \quad & 1X = 0 \\ & X \in \mathcal{K} := \mathbb{R} \end{aligned} \tag{137}$$

and its dual problem

$$\begin{aligned} \underset{Y\in\mathbb{R},\ y\in\mathbb{R}}{\text{maximize}} \quad & 0y \\ \text{subject to} \quad & 1y - Y = 0 \\ & Y \in \mathcal{K}^P = \{0\} \end{aligned} \tag{138}$$

for which we obtain analogously to our first example $\mathcal{H} = 1-2(1(\frac{1}{1^2})1) = -1$ and a function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(Z) = \tfrac{Z\mathcal{H}Z}{4} - \tfrac{Z|Z|_{\mathbb{R}}}{4} = \tfrac{-Z^2}{2} \tag{139}$$

with deriative $\nabla f(Z) = -Z = -\underbrace{\mathcal{H}}_{=-1}[\nabla f(Z)]$. Obviously $Z^{(*)} := 0$ is its unique zero. Let $Z^{(0)} \in E$ be arbitrary. If we rewrite the iterations

$$Z^{(k)} = Z^{(k-1)} - 2t\mathcal{H}[\nabla f(Z^{(k-1)})] = (1-2t)Z^{(k-1)} = (1-2t)^{(k)}Z^{(0)} \tag{140}$$

we obtain

$$||\nabla f(Z^{(k)})||_2^2 = (2t-1)^{2k}||Z^{(0)} - 0||_2^2 = (2t-1)^{2k}||Z^{(0)} - Z^{(*)}||_2^2 \tag{141}$$

which proves our claim. ■

Now the proof above did not really explain our reasoning for choosing these specific (and essentially trivial) examples. In fact there is more to it, than the eye can see: They are chosen as reverse-engineered solutions of a specific SDP-Relaxation. Namely a relaxation of the worst-case-performance after κ steps problem. The idea behind this comes from worst-case-performance-estimation for unconstrained convex optimization problems (derived in the ground-breaking work in [24],[83]) and is transfered to the broader setting of fixed point methods in this thesis: One tries to express all the inequalities from section 2 in terms of a special Gram-Matrix and attempts to maximize $\left\|\nabla f(Z^{k)})\right\|$ over all functions f by relaxing the Gram-Matrix to a positive semidefinite Matrix. While the tightness of the approach in [83] is guaranteed (via a convex interpolation theorem), it it is not guaranteed in our situation (i.e. for $\mathcal{H}\nabla f$), but luckily the reverse-engineered solutions fit our setting. The main difference between our approach and the approach in [83], is that we do not make any assumptions regarding convexity: This leads to a different angle, namely worst-case-performance-estimation for finding fixed points. The following corollary anticipates the main result from the next subsection:

Corollary 4.7 (Worst-case-complexity (tightness of complexity for constant step length part 3))**.** *Let $\kappa \in \mathbb{N}$, $t \in [\frac{1}{2}, 1)$, $\nabla f : E \to E$ defined as in* (18) *and $Z^{(*)} \in E$ with $\nabla f(Z^{(*)}) = 0$. For the iterates defined by $Z^{(k+1)} := Z^{(k)} - 2t\mathcal{H}[\nabla f(Z^{(k)})]$ the following inequality always holds true:*

$$\left\|\nabla f(Z^{(\kappa)})\right\|^2 \leq \begin{cases} \frac{1}{\kappa+1}(\frac{\kappa}{\kappa+1})^\kappa \frac{\left\|Z^{(0)}-Z^{(*)}\right\|^2}{4t(1-t)} & \textit{if } \frac{1}{2} \leq t \leq \frac{1}{2}(1+\sqrt{\frac{\kappa}{\kappa+1}}) \\ (2t-1)^{2\kappa}\left\|Z^{(0)} - Z^{(*)}\right\|^2 & \textit{if } \frac{1}{2}(1+\sqrt{\frac{\kappa}{\kappa+1}}) < t < 1 \end{cases} \tag{142}$$

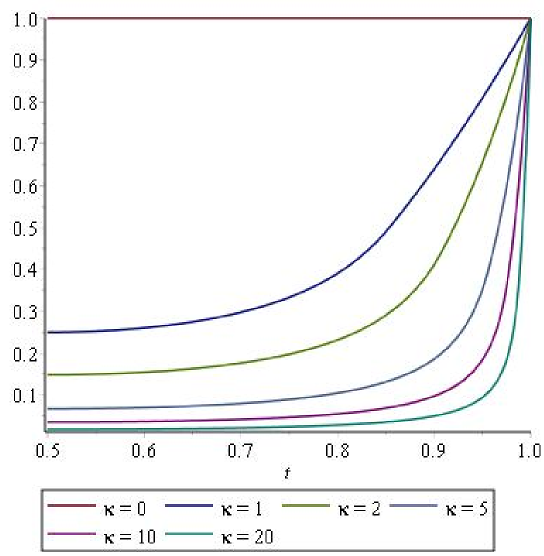


Figure 3: Right hand side of (142) for $\left\|Z^{(0)} - Z^{(*)}\right\| = 1$ and different values of κ

Proof. Let $T : E \to E$ be defined via

$$T(Z) := Z - 2\mathcal{H}\nabla f(Z) = \mathcal{H}[|Z|_\mathcal{K} - 2\mathcal{R}] \tag{143}$$

then T is nonexpansive (i.e. 1-Lipschitz continuous) and every $Z^{(*)} \in E$ with $\nabla f(Z^{(*)}) = 0$ is a fixed point of T, i.e. satisfies $T(Z^{(*)}) = Z^{(*)}$. Furthermore we can rewrite the iteration $Z^{(k+1)} = Z^{(k)} - 2t\mathcal{H}[\nabla f(Z^{(k)})] = (1-t)Z^{(k)} + tT(Z^{(k)})$. Our corollary then follows from theorem 4.9 and bound (146) below, which is proved for (possibly infinite dimensional) real Hilbert spaces in section 4.1.

■

Let us close this subsection with pointing out that the bound (142) in corollary 4.7 above is optimal (i.e. one can not improve the convergence rate without making further assumptions), since a better rate would contradict our examples from theorems 4.4 and 4.6 (note the equality Now one might be disappointed by this insight, since we do know other methods with much better worst case convergence rates (and will briefly discuss one in section 4.2). However there is no need to disregard our method, especially since any rate of $\left\|\nabla f(Z^{(k)})\right\|$ is in most cases only a vehicle to find the optimal value of (1). If we recall the rough inequality (41), it is easily seen from monotonicity that

$$|f(Z^{(*)}) - f(Z^{(k)})| \leq \left\|\nabla f(Z^{(k)})\right\| \left\|Z^{(k)} - Z^{(*)}\right\| \leq \left\|\nabla f(Z^{(k)})\right\| \left\|Z^{(0)} - Z^{(*)}\right\|$$

holds and therefore the rate from our corollary readily implies $|f(Z^{(*)}) - f(Z^{(k)})| \in O(k^{-\frac{1}{2}})$. However we can do significantly better by using the finer inequalities (45) and by averaging over the function values $f(Z^{(k)})$ as the next proposition shows.

Proposition 4.8. *Let $\kappa \in \mathbb{N}_0$, $t \in [\frac{1}{2}, 1)$, $f : E \to \mathbb{R}$ defined as in (25) and $Z^{(*)} \in E$ with $\nabla f(Z^{(*)}) = 0$. For the κ-th iterate of the sequence $Z^{(k+1)} = Z^{(k)} - 2t\mathcal{H}[\nabla f(Z^{(k)})]$ the following inequality always holds true:*

$$|f(Z^{(*)}) - \frac{1}{\kappa+1}\sum_{k=0}^{\kappa} f(Z^{(k)})| \leq \frac{1}{\kappa+1}\frac{\left\|Z^{(0)} - Z^{(*)}\right\|^2}{8t(1-t)} \tag{144}$$

The proof has been moved to the Appendix, since it is slightly longer than appropriate at this point and does not directly fit into our upcoming analysis. Tackling this proof is encouraged only after reading the next section, where underlying concepts are introduced.

4.1 Worst-Case-Complexity of Krasnoselski-Mann Iteration

For the rest of this section we will shift our focus from finding zeros to finding fixed points. We will also loosen our setting a bit by not assuming finite dimensionality. Let $\mathbb{H}$ be a Hilbert space equipped with a symmetric inner product $\langle .,. \rangle : \mathbb{H} \times \mathbb{H} \to \mathbb{R}$. Let $T : \mathbb{H} \to \mathbb{H}$ be a nonexpansive mapping and consider for fixed $x_0 \in \mathbb{H}$ the Krasnoselski-Mann iteration (KM iteration), see [43] and [57],

$$x_{k+1} := (1-t_k)x_k + t_k T(x_k) \tag{145}$$

with $t_k \in (0,1) \forall k \in \mathbb{N}_0$ for approximating a fixed point of T. Let $\|x\| = \sqrt{\langle x, x \rangle}$ denote the induced norm and $Fix(T) := \{x \in H \mid x = T(x)\}$ the set of fixed points of T. It is well known (and we also showed it in the previous section with different notation) that if the set $Fix(T)$ is nonempty and $\sum_{k=1}^{\infty} t_k(1-t_k) = \infty$, then the sequence of the norm of residuals $\|x_k - T(x_k)\|$ tends to zero, i.e. $\lim_{k\to\infty} \|x_k - T(x_k)\| = 0$. Our goal here is to quantify their rate of convergence for constant step length $t_k \equiv t \in [\frac{1}{2}, 1]$ in an optimal fashion. Regarding related results concerning the case of Hilbert and normed spaces (and non-constant t_k), I would especially like to mention [90], [14], [46] and [47], which were certainly quite relevant for my analysis. Here, to the best of my knowledge, all existing results are substantially improved for the setting of (real) Hilbert spaces. The proof is based on semidefinite programming and will be generalized in the next section. The proof got strongly motivated by the recent work of Taylor et al. [83] on worst case performance of first order unconstrained minimization methods. However its methodology is slightly different and the focus is entirely different: The technique in [83] is meant for analyzing first order methods for minimizing smooth convex functions, such as gradient method (GM) or the famous method due to Nesterov [65]. Here we do not have a function to minimize and in general T is not a gradient mapping[11]. At this point I would like to thank Adrien Taylor and François Glineur, who I had the pleasure of first meeting on the ISMP conference held in Pittsburgh in 2015. After attending François's talk, we had a discussion regarding my particular application. Although we came to no substantial conclusion, it seemed that their technique could not be immediately transfered, as it relies on convex interpolation (and the reduced Lagrangian is (usually) not convex). The main complication at the time was thinking of iteration (109) as a "mirrored" gradient method, rather than a fixed point method (KM iteration), which as we have seen from the examples of the previous section, does not lead to better convergence results. The paradigm shift from gradient based methods towards fixed point methods, then finally led to the upcoming analysis presented next. Let us stretch the fact, that our previous analysis shows that the following bound can not be improved without further assumptions.

Theorem 4.9. *Let $(\mathbb{H}, \langle ., . \rangle)$ be a real Hilbert space, let $T : \mathbb{H} \to \mathbb{H}$ be nonexpansive and let $x_0 \in \mathbb{H}$ be arbitrary but fixed. If T has fixed points, i.e. $Fix(T) \neq \emptyset$, then the KM iterates defined in* (145) *for $t_k \equiv t \in [\frac{1}{2}, 1]$ $\forall k \in \mathbb{N}_0$ satisfy*

$$\left\| \tfrac{1}{2}(x_k - T(x_k)) \right\|^2 \leq \begin{cases} \frac{1}{k+1}\left(\frac{k}{k+1}\right)^k \frac{\|x_0 - x_*\|^2}{4t(1-t)} & \text{if } \frac{1}{2} \leq t \leq \frac{1}{2}\left(1 + \sqrt{\frac{k}{k+1}}\right) \\ (2t-1)^{2k} \|x_0 - x_*\|^2 & \text{if } \frac{1}{2}\left(1 + \sqrt{\frac{k}{k+1}}\right) < t \leq 1 \end{cases} \tag{146}$$

$\forall k \in \mathbb{N}_0$ $\forall x_* \in Fix(T)$. This bound is tight.

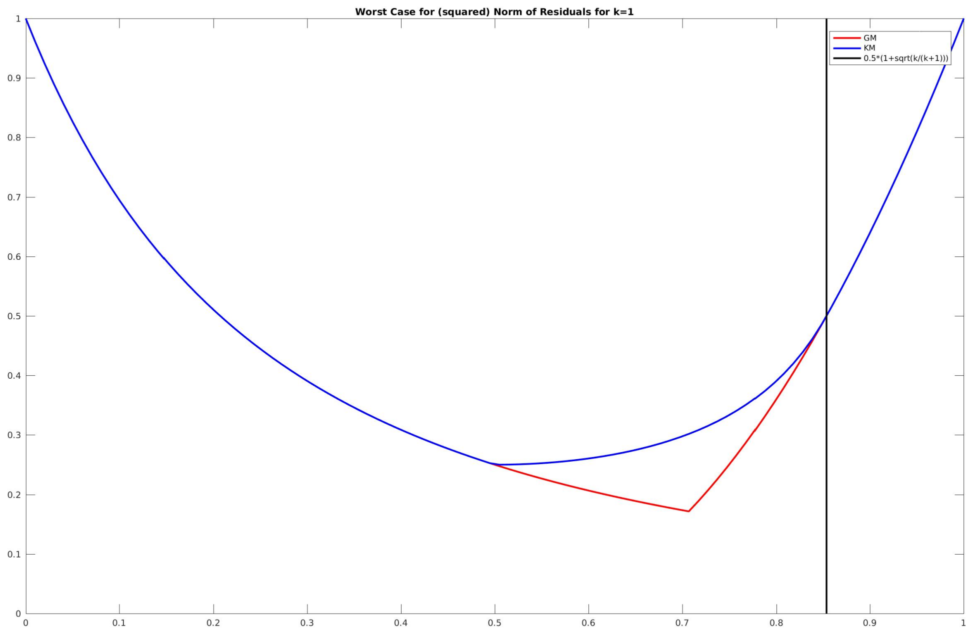
Remark 4.10. *The case of a fixed step length in the interval $[0, \frac{1}{2})$ in the above theorem is excluded, since all numerical evaluations suggested that such a small step length is inferior to larger ones. However, as we shall see below, for fixed k we can actually plot the worst case with respect to t. Matching these plots to conjecture 3 from [83] (regarding the worst case complexity of gradients method for smooth convex functions) leads us to believe, that*

$$\left\| \tfrac{1}{2}(x_k - T(x_k)) \right\|^2 \leq \begin{cases} \frac{\|x_0 - x_*\|^2}{(2kt+1)^2} & \text{if } 0 \leq t \leq c_k \\ \frac{1}{k+1}\left(\frac{k}{k+1}\right)^k \frac{\|x_0 - x_*\|^2}{4t(1-t)} & \text{if } c_k \leq t \leq \frac{1}{2} \end{cases} \tag{147}$$

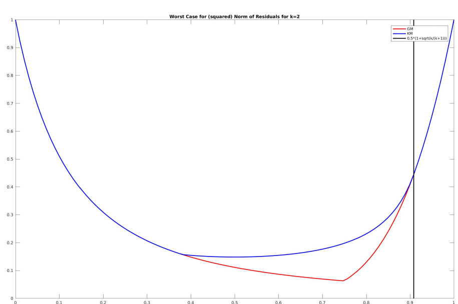
[11] However if T is a gradient mapping, then the residual mapping $g(x) := \frac{1}{2}(x - T(x))$ is the gradient of a smooth convex function. This implies that the class considered here, is strictly larger than the one considered in [83].

is a tight worst case rate for $t \in [0, \frac{1}{2})$, where $c_k \in (0, \frac{1}{2}]$ is one of the two points where the two components on the right hand side of (147) intersect given by

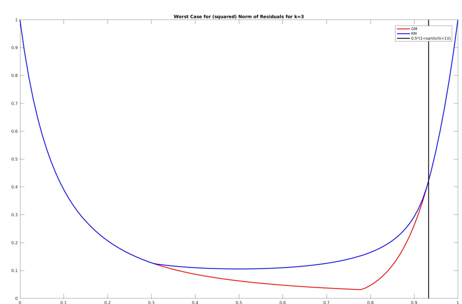
$$c_k := \frac{1}{2} \frac{(k+1) - k(\frac{k}{k+1})^k + \sqrt{(k+1)^2 - (2k^2+3k+1)(\frac{k}{k+1})^k}}{k^2(\frac{k}{k+1})^k + k + 1} \tag{148}$$



(a) $k = 1$ and $||x_0 - x_*|| = 1$



(b) $k = 2$ and $||x_0 - x_*|| = 1$



(c) $k = 3$ and $||x_0 - x_*|| = 1$

Figure 4: MATLAB plot for different values of k with respect to t on the interval $[0, 1]$, red line marks conjecture 3 from [83], blue line marks our worst-case plot. We will see that the blue line presents the worst-case correctly on the interval $[\frac{1}{2}, 1]$ and coincides with the right hand side of (146).

We will partially prove conjecture 3 of Taylor et al. (on the interval $[\frac{1}{2}(1+\sqrt{\frac{k}{k+1}}), \leq 1]$*) as the gradient method is equivalent to the KM-iteration for smooth convex functions.*

Proof of Theorem 4.9: Let $x_* \in Fix(T)$. The KM iteration was stated in the form (145) to comply with existing literature. For our proof however, it is more convenient to consider the shifted sequence $\bar{x}_1 := x_0$ and $\bar{x}_k := x_{k-1}$ $\forall k \in \mathbb{N}_{\neq 0}$ and to show $\forall k \in \mathbb{N}$ $\forall x_* \in Fix(T)$ a shifted statement

$$\left\|\tfrac{1}{2}(\bar{x}_k - T(\bar{x}_k))\right\|^2 \leq \begin{cases} \frac{1}{k}(\frac{k-1}{k})^{k-1}\frac{\|\bar{x}_1 - x_*\|^2}{4t(1-t)} & \text{if } \frac{1}{2} \leq t \leq \frac{1}{2}(1+\sqrt{\frac{k-1}{k}}) \\ (2t-1)^{2(k-1)}\|\bar{x}_1 - x_*\|^2 & \text{if } \frac{1}{2}(1+\sqrt{\frac{k-1}{k}}) < t \leq 1 \end{cases} \tag{149}$$

Let us define $g(x) := \frac{1}{2}(x - T(x))$. It is well known that g is firmly nonexpansive. For sake of completeness the argument is repeated here:

$$\begin{aligned} &\|g(x)-g(y)\|^2 - \langle g(x)-g(y), x-y\rangle \\ =&\left\|g(x)-g(y)-\tfrac{1}{2}(x-y)\right\|^2 - \tfrac{1}{4}\|x-y\|^2 \\ =&\tfrac{1}{4}\|T(x)-T(y)\|^2 - \tfrac{1}{4}\|x-y\|^2 \leq 0 \ \forall x,y \in H. \end{aligned}$$

Nonexpansiveness and the Cauchy-Schwarz inequality imply $\|g(x)-g(y)\| \leq \|x-y\|$ $\forall x,y \in H$. For $k=1$ the statement (149) follows immediately since $g(x_*)=0$ and therefore $\frac{1}{2}\|\bar{x}_1 - T(\bar{x}_1)\| = \|g(\bar{x}_1)\| = \|g(\bar{x}_1) - g(\bar{x}_*)\| \leq \frac{\|\bar{x}_1 - x_*\|}{1}$.
For fixed $k \geq 2$ we first consider the differences $\bar{x}_j - \bar{x}_1$ for $j \in \{2,..,k\}$

$$\begin{aligned} \bar{x}_j - \bar{x}_1 &= x_{j-1} - \bar{x}_1 \\ &= x_{j-2} - t(x_{j-2} - T(x_{j-2})) - \bar{x}_1 \\ &= x_{j-2} - 2tg(x_{j-2}) - \bar{x}_1 \\ &= \bar{x}_{j-1} - 2tg(\bar{x}_{j-1}) - \bar{x}_1 \end{aligned}$$

which inductively leads to

$$\bar{x}_j - \bar{x}_1 = -2t\sum_{l=1}^{j-1} g(\bar{x}_l).$$

Let us shorten the notation and define $g_i := g(\bar{x}_i)$, $R := \|\bar{x}_1 - x_*\| \geq 0$, the vector $b = (\langle g_i, \bar{x}_1 - x_*\rangle)_{i=1}^k$, the matrices $A := (\langle g_i, g_j\rangle)_{i,j=1}^k$ and

$$L := -2t\begin{pmatrix} 0 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{k \times k}.$$

Note that $\begin{pmatrix} R^2 & b^T \\ b & A \end{pmatrix} \in \mathbb{R}^{(k+1)\times(k+1)}$ is a Gramian matrix formed from $\bar{x}_1 - x_*, g_1, \dots, g_k \in \mathbb{H}$ and is therefore symmetric and positive semidefinite. We proceed by expressing the inequalities from firm nonexpansiveness in terms of the Gramian matrix. Since L often is of much lower dimension than $\mathbb{H}$, this is sometimes referred to as 'Kernel-Trick'. Keeping in mind that we can write $\bar{x}_j - \bar{x}_1 = -2t\sum_{l=1}^{j-1} g_l$ for $j \in \{1,..,k\}$ we arrive at

$$AL = ((\langle g_i, \bar{x}_j - \bar{x}_1\rangle))_{i,j=1}^k.$$

Let $e \in \mathbb{R}^k$ denote the vector of all ones. Then

$$\text{diag}(AL)e^T - AL = ((\langle g_i, \bar{x}_i - \bar{x}_j\rangle))_{i,j=1}^k.$$

where diag(.) denotes the diagonal of its (square) matrix argument. Let L^T, e^T denote the transpose of L respectively e. Hence

$$\text{diag}(AL)e^T + e\ \ \text{diag}(AL)^T - AL - L^T A = ((\langle g_i - g_j, \bar{x}_i - \bar{x}_j\rangle))_{i,j=1}^k$$

and

$$be^T + AL = ((\langle g_i, \bar{x}_j - x_*\rangle))_{i,j=1}^k,$$

$$\text{diag}(A)e^T + e\ \ \text{diag}(A)^T - 2A = (\|g_i - g_j\|^2)_{i,j=1}^k.$$

The firm nonexpansiveness inequalities $\|g_i - g_j\|^2 \leq \langle g_i - g_j, \bar{x}_i - \bar{x}_j\rangle$ are equivalent to the component-wise inequality

$$\text{diag}(A)e^T + e\ \ \text{diag}(A)^T - 2A \leq \text{diag}(AL)e^T + e\ \ \text{diag}(AL)^T - AL - L^T A. \tag{150}$$

Note that only $\frac{k^2-k}{2}$ of these componentwise inequalities are non redundant. From $g_* := g(x_*) = 0$ we obtain another k inequalities i.e. $\|g_i\|^2 \leq \langle g_i, \bar{x}_i - x_*\rangle$ which translate to

$$\text{diag}(A) \leq b + \text{diag}(AL). \tag{151}$$

Defining $U := I - L$, relations (150) and (151) can be shortened slightly to

$$\text{diag}(AU)e^T + e\ \ \text{diag}(AU)^T \leq AU + U^T A$$

and

$$\text{diag}(AU) \leq b.$$

Let $e_k \in \mathbb{R}^k$ denote the k-th unit vector, recall that $\mathbb{S}^n := \{X \in \mathbb{R}^{n\times n} \mid X = X^T\}$ denotes the space of symmetric matrices and $\mathbb{S}^n_+ := \{X \in \mathbb{S}^n \mid x^T X x \geq 0\ \forall x \in \mathbb{R}^n\}$ the convex cone of positive semidefinite matrices. Consider the chain of inequalities

$$\begin{aligned}\|g(\bar{x}_k)\|^2 = \underset{y_0\in\mathbb{R}, y_1\in\mathbb{R}^k, Y_2\in\mathbb{S}^k}{\text{maximize}} (Y_2)_{kk} \mid & \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2\end{pmatrix} \in \mathbb{S}^{k+1}_+,\ y_0 \leq R^2,\ diag(Y_2 U) \leq y_1 \\ & \mid \text{diag}(Y_2 U)e^T + e\ \ \text{diag}(U^T Y_2)^T \leq Y_2 U + U^T Y_2 \\ & \mid y_0 = R^2, y_1 = b, Y_2 = A\end{aligned}$$

$$\begin{aligned}\leq \underset{y_0\in\mathbb{R}, y_1\in\mathbb{R}^k, Y_2=Y_2^T\in\mathbb{S}^k}{\text{maximize}} (Y_2)_{kk} \mid & \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2\end{pmatrix} \in \mathbb{S}^{k+1}_+,\ y_0 \leq R^2,\ \text{diag}(Y_2 U) \leq y_1 \\ & \mid \text{diag}(Y_2 U)e^T + e\ \ \text{diag}(U^T Y_2)^T \leq Y_2 U + U^T Y_2\end{aligned} \tag{152}$$

$$\leq \underset{\xi\in\mathbb{R}_+, X\in\mathbb{S}^k\cap\mathbb{R}_+^{k\times k}}{\text{minimize}} \quad R^2\xi \mid \begin{pmatrix} \xi & -\frac{1}{2}\operatorname{diag}(X)^T \\ -\frac{1}{2}\operatorname{diag}(X) & UF(X)+F(X)U^T \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & e_k e_k^T \end{pmatrix} \in \mathbb{S}_+^{k+1} \tag{153}$$

for $F(X) := \operatorname{Diag}(Xe) + \frac{1}{2}\operatorname{Diag}(\operatorname{diag}(X)) - X$. The first equality follows from construction, the first inequality from relaxing, and the second inequality from weak conic duality, which we shall proof first: Define the Euclidean space $\mathbb{E} := \mathbb{R}\times\mathbb{S}^k\times\mathbb{S}^{k+1}$ and the self-dual closed convex cone $\mathcal{K} := \mathbb{R}_+ \times (\mathbb{S}^k\cap\mathbb{R}_+^{k\times k}) \times \mathbb{S}_+^{k+1}$. Define $C = (R^2, 0, 0)^* \in \mathbb{E}$ and $\tilde{b} = e_{\frac{(k+1)(k+2)}{2}}$. The problem (152) can then be written as a conic optimization problem in dual standard form, i.e.

$$\max\{\tilde{b}^T\lambda \mid \mathcal{A}^*(\lambda) + S = C,\ S\in\mathcal{K},\ \lambda\in\mathbb{R}^{\frac{(k+2)(k+1)}{2}}\}$$

where $\mathcal{A}^* : \mathbb{R}^{\frac{(k+2)(k+1)}{2}} \to \mathbb{E}$ is given as a composition $\mathcal{A}^* = \bar{\mathcal{A}}^*(smat(\lambda))$ for $\bar{\mathcal{A}}^* : \mathbb{S}^{k+1}\to\mathbb{E}$ given by

$$\bar{\mathcal{A}}^*(\begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix}) = \begin{pmatrix} y_0 \\ \operatorname{diag}(Y_2U)e^T + e\ \operatorname{diag}(U^TY_2)^T - (Y_2U + U^TY_2) + \operatorname{Diag}(\operatorname{diag}(Y_2U) - y_1) \\ -\begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \end{pmatrix}.$$

(Note that this is only true because $\operatorname{diag}(Y_2U)e^T + e\ \operatorname{diag}(U^TY_2)^T - (Y_2U + U^TY_2)$ has an all zero diagonal.) The dual of this problem (i.e. the primal standard form) then takes the form

$$\min\{\langle C, \widetilde{X}\rangle_{\mathbb{E}} \mid (\mathcal{A}^*)^*(\widetilde{X}) = \tilde{b}\ \widetilde{X}\in\mathbb{E}, \widetilde{X}\in\mathcal{K}\}$$

where $(\mathcal{A}^*)^* : \mathbb{E}\to\mathbb{R}^{\frac{(k+2)(k+1)}{2}}$ denotes the adjoint operator of $\mathcal{A}^*$. For $\widetilde{X} = \begin{pmatrix} \xi \\ X_1 \\ X_2 \end{pmatrix} \in \mathbb{E}$ we have the explicit expression

$$\operatorname{smat}((\mathcal{A}^*)^*(\widetilde{X})) = (\bar{\mathcal{A}}^*)^*(\widetilde{X}) = \begin{pmatrix} \xi & -\frac{1}{2}\operatorname{diag}(X_1)^T \\ -\frac{1}{2}\operatorname{diag}(X_1) & UF(X_1)+F(X_1)U^T \end{pmatrix} - X_2$$

as well as $smat(\tilde{b}) = \begin{pmatrix} 0 & 0 \\ 0 & e_ke_k^T \end{pmatrix}$ and $\langle C, \widetilde{X}\rangle_{\mathbb{E}} = R^2\xi$. By eliminating the variable $X_2\in\mathbb{S}_+^{k+1}$ we obtain the claimed form (153).

We conclude the proof by showing feasibility of certain $\hat{\xi}$ and $\hat{X}$ for (153), which were obtained by reverse engineering from numerical solution(s) of (153) for small values of k. We note that even for fixed k and t (153) may have infinitely many optimal solutions. Luckily however, we can construct tridiagonal solutions, which are much easier to handle as they seem to be unique. Proving feasibility of these solutions involves a number of matrix operations, which can in principle be done "by hand". This cumbersome task has been performed and afterwards verified by using the computer algebra program MAPLE. Since most of these calculations get rather long, only the main steps and results are presented here. From this all intermediate steps are in principle reproducible. We distinguish two cases:

Case 1: In our first case we will assume $k\geq 2$ as well as $\frac{1}{2}\leq t\leq\frac{1}{2}(1+\sqrt{1-\frac{1}{k}})$. Tridiagonal matrices will be denoted by $Tridiag(.,.,.)$. Let us define the real number $\hat{\xi} = \frac{1}{4t(1-t)k}(\frac{k-1}{k})^{k-1}$ the componentwise nonnegative vectors $x_D\in\mathbb{R}_+^k$, $x_L\in\mathbb{R}_+^{k-1}$ and the symmetric matrix $\hat{X}\in\mathbb{R}_+^{k\times k}$:

$$
\begin{aligned}
(\hat{x}_D)_j &:= \frac{(k+1-j)t-1}{k}(\frac{k-1}{k})^{k-1-j} \quad \text{for } j \in \{1,..,k-1\} \\
(\hat{x}_D)_k &:= (1-t) \\
(\hat{x}_L)_j &:= \frac{j(1-t)}{2}(\frac{k-1}{k})^{k-1-j} \text{ for } j \in \{1,..,k-1\} \\
\hat{X} &:= \frac{1}{k(1-t)t} Tridiag(\hat{x}_L, \hat{x}_D, \hat{x}_L) \\
&= \frac{1}{k(1-t)t} \begin{pmatrix} \ddots & & \ddots & & \\ \ddots & \frac{(k+1-j)t-1}{k}(\frac{k-1}{k})^{k-1-j} & \frac{i(1-t)}{2}(\frac{k-1}{k})^{k-1-i} & & \\ & \frac{j(1-t)}{2}(\frac{k-1}{k})^{k-1-j} & \ddots & \ddots & \\ & & \ddots & (1-t) \end{pmatrix}.
\end{aligned} \tag{154}
$$

Defining the k dimensional vector

$$
\begin{aligned}
(\hat{F}_D)_j &:= \frac{(j(2(1-t)-\frac{1}{k})+2t-1)}{2}(\frac{k-1}{k})^{k-1-j} \quad \text{for } j \in \{1,..,k-1\} \\
(\hat{F}_D)_k &:= \frac{k(1-t)}{2}
\end{aligned} \tag{155}
$$

leads to the straightforward but tedious evaluation

$$
\begin{aligned}
F(\hat{X}) &= \operatorname{Diag}(\hat{X}e) + \tfrac{1}{2}\operatorname{Diag}(\operatorname{diag}(\hat{X})) - \hat{X} \\
&= \frac{1}{k(1-t)t} Tridiag(-\hat{x}_L, \hat{F}_D, -\hat{x}_L) \\
&= \frac{1}{k(1-t)t} \begin{pmatrix} \ddots & & \ddots & & \\ \ddots & \frac{(j(2(1-t)-\frac{1}{k})+2t-1)}{2}(\frac{k-1}{k})^{k-1-j} & -\frac{i(1-t)}{2}(\frac{k-1}{k})^{k-1-i} & & \\ & -\frac{j(1-t)}{2}(\frac{k-1}{k})^{k-1-j} & \ddots & \ddots & \\ & & \ddots & \frac{k(1-t)}{2} \end{pmatrix}.
\end{aligned} \tag{156}
$$

Now to prove the result in our first case, it suffices to show that the matrix inequality in (153) holds true for the chosen values $\hat{\xi}$ and $\hat{X}$. In order to do so, we first apply a congruence transformation, which converts the constraint of (153) to a tridiagonal matrix inequality for the given values $\hat{\xi}$ and $\hat{X}$. For this purpose we define a vector $a \in \mathbb{R}^k$ with entries $a_1 := 0,\ a_2 := \frac{-t}{k(1-t)}, a_j := a_{j-1}(2 - \frac{1}{k(1-t)}) - a_{j-2}\frac{k-1}{k}\ \forall 3 \leq j \leq k$ and the Toeplitz matrix

$$
\hat{U} = \begin{pmatrix} 1 & a_1 & a_2 & \dots & a_{k-1} \\ 0 & 1 & a_1 & \dots & a_{k-2} \\ \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & \vdots & 0 & 1 & a_1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{157}
$$

The vector $a \in \mathbb{R}^k$ has been chosen to satisfy $(UF(\hat{X}) + F(\hat{X})U^T - e_k e_k^T)a = -t \ \operatorname{diag}(\hat{X})$ and $a^T \operatorname{diag}(\hat{X}) = -4t\hat{\xi}$. This implies

$$\begin{aligned}
&\begin{pmatrix} -2t & a^T \\ 0 & \hat{U} \end{pmatrix} \begin{pmatrix} \hat{\xi} & -\frac{1}{2}\operatorname{diag}(\hat{X})^T \\ -\frac{1}{2}\operatorname{diag}(\hat{X}) & UF(\hat{X}) + F(\hat{X})U^T - e_k e_k^T \end{pmatrix} \begin{pmatrix} -2t & 0 \\ a & \hat{U}^T \end{pmatrix} \\
&= \begin{pmatrix} -2t\hat{\xi} - \frac{1}{2}\operatorname{diag}(\hat{X}) & t\operatorname{diag}(\hat{X})^T + a^T(UF(\hat{X}) + F(\hat{X})U^T - e_k e_k^T) \\ -\frac{1}{2}\hat{U}\operatorname{diag}(\hat{X}) & \hat{U}(UF(\hat{X}) + F(\hat{X})U^T - e_k e_k^T) \end{pmatrix} \begin{pmatrix} -2t & 0 \\ a & \hat{U}^T \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ -\frac{1}{2}\hat{U}\operatorname{diag}(\hat{X}) & \hat{U}(UF(\hat{X}) + F(\hat{X})U^T - e_k e_k^T) \end{pmatrix} \begin{pmatrix} -2t & 0 \\ a & \hat{U}^T \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 \\ 0 & \hat{U}(UF(\hat{X}) + F(\hat{X})U^T - e_k e_k^T)\hat{U}^T \end{pmatrix}
\end{aligned} \tag{158}$$

with a tridiagonal matrix

$$\begin{aligned}
M :=& \hat{U}(UF(\hat{X}) + F(\hat{X})U^T - e_k e_k^T)\hat{U}^T \\
=& \begin{pmatrix} \ddots & & \ddots & & \\ \ddots & \frac{j(k-1)^{k-2-j}}{k^{k-j+1}} \frac{(kt-(k-1))^2}{t(1-t)} + \frac{(j-1)(k-1)^{k-1-j}}{k^{k-j}} \frac{1-t}{t} & \frac{i(k-1)^{k-2-i}}{k^{k-i}} \frac{kt-(k-1)}{t} & & \\ & \frac{j(k-1)^{k-2-j}}{k^{k-j}} \frac{kt-(k-1)}{t} & \ddots & \ddots & \\ & & \ddots & \frac{(1-t)}{t} \end{pmatrix}
\end{aligned} \tag{159}$$

where i denotes the corresponding row and j the corresponding column. We identify M as positive semidefinite by using the Schur-Complement: Define $m_j := \frac{j(k-1)^{k-2-j}}{k^{k-j+1}} \frac{(kt-(k-1))^2}{t(1-t)} > 0$ for $j \in \{1, .., k\}$. Note that

$$M_{kk} - \frac{M_{k,k-1}^2}{m_{k-1}} = 0 \tag{160}$$

and therefore the 2x2 matrix

$$\begin{pmatrix} m_{k-1} & M_{(k-1)k} \\ M_{k(k-1)} & M_{kk} \end{pmatrix} \succeq 0 \tag{161}$$

is positive semidefinite. By inductively using the equalities

$$M_{j+1,j+1} - \frac{M_{j+1,j}^2}{m_j} = m_{j+1} > 0 \tag{162}$$

for $j \in \{1, .., k-2\}$ we see that the matrices

$$\begin{pmatrix} m_j & M_{j(j+1)} & 0^T \\ M_{(j+1)j} & \multicolumn{2}{c}{M_{j+1:k,j+1:k}} \\ 0 & & \end{pmatrix} \succeq 0 \tag{163}$$

are also positive semidefinite and since

$$M_{11} = \frac{(k-1)^{k-3}}{k^k} \frac{(kt-(k-1))^2}{t(1-t)} = m_1 > 0 \tag{164}$$

is positive, this concludes the first case of our proof.

Case 2: In our second case we will assume $k \geq 2$ and $1 \geq t > \frac{1}{2}(1+\sqrt{1-\frac{1}{k}})$. Again we define the real number $\hat{\xi} = (2t-1)^{2k-2}$ and the componentwise nonnegative vectors $x_D \in \mathbb{R}^k_+$, $x_L \in \mathbb{R}^{k-1}_+$ and the matrix $\hat{X} \in \mathbb{R}^{k\times k}_+$ (Note that the variables have changed in comparison to the first part):

$$\begin{aligned}
(\hat{x}_D)_j &:= (4t(2t-1) - \tfrac{4(j-1)t+2}{k})(2t-1)^{2(k-1-j)} \quad \text{for } j \in \{1,..,k-1\}\\
(\hat{x}_D)_k &:= 2 - \tfrac{2}{2t-1}\tfrac{k-1}{k}\\
(\hat{x}_L)_j &:= \tfrac{j}{2tk}(2t-1)^{2(k-1-j)} \text{ for } j \in \{1,..,k-1\}\\
\hat{X} &:= Tridiag(\hat{x}_L, \hat{x}_D, \hat{x}_L)\\
&= \begin{pmatrix} \ddots & & \ddots & & \\ \ddots & (4t(2t-1) - \frac{4(j-1)t+2}{k})(2t-1)^{2(k-1-j)} & \frac{i}{2tk}(2t-1)^{2(k-1-i)} & & \\ & \frac{j}{2tk}(2t-1)^{2(k-1-j)} & \ddots & & \ddots \\ & & \ddots & & 2-\frac{2}{2t-1}\frac{k-1}{k} \end{pmatrix}.
\end{aligned} \tag{165}$$

Defining the vector (which again has changed in comparison with the first part)

$$\begin{aligned}
(\hat{F}_D)_j &: = (2t-1)^{2(k-1-j)}(2t-1)(2t - \frac{2j-1}{2tk}) \quad \text{for } j \in \{1,..,k-1\}\\
(\hat{F}_D)_k &: = 1 - \frac{1}{(2t-1)2t}\frac{k-1}{k}
\end{aligned} \tag{166}$$

leads to the straightforward evaluation

$$\begin{aligned}
F(\hat{X}) &= Diag(\hat{X}e) + \tfrac{1}{2}\operatorname{Diag}(\operatorname{diag}(\hat{X})) - \hat{X}\\
&= Tridiag(-\hat{x}_L, \hat{F}_D, -\hat{x}_L)\\
&= \begin{pmatrix} \ddots & & \ddots & & \\ \ddots & (2t-1)^{2(k-1-j)}(2t-1)(2t-\frac{2j-1}{2tk}) & -\frac{i(1-t)}{2} - \frac{i}{2tk}(2t-1)^{2(k-1-i)} & & \\ & -\frac{j}{2tk}(2t-1)^{2(k-1-j)} & \ddots & & \ddots \\ & & \ddots & & 1-\frac{1}{(2t-1)2t}\frac{k-1}{k} \end{pmatrix}.
\end{aligned} \tag{167}$$

We now proceed similarly to the first case by applying a congruence transformation to the matrix inequality (153). Here the transformation is slightly simpler and depends only on the inverse of $U = I - L$ given by

$$U^{-1} = \begin{pmatrix} 1 & -2t(1-2t)^0 & -2t(1-2t)^1 & \dots & -2t(1-2t)^{k-2}\\ 0 & 1 & -2t(1-2t)^0 & \dots & -2t(1-2t)^{k-3}\\ \vdots & 0 & \ddots & \ddots & \vdots\\ 0 & \vdots & 0 & 1 & -2t(1-2t)^0\\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{168}$$

Another straightforward computation reveals $(UF(\hat{X})+F(\hat{X})U^T-e_ke_k^T)U^{-T}e=\frac{1}{2}\operatorname{diag}(X)$ and $e^TU^{-1}\operatorname{diag}(\hat{X})=2\hat{\xi}$ which implies

$$\begin{aligned}
&\begin{pmatrix}-2t & -2te^TU^{-1}\\ 0 & U^{-1}\end{pmatrix}\begin{pmatrix}\hat{\xi} & -\frac{1}{2}\operatorname{diag}(\hat{X})^T\\ -\frac{1}{2}\operatorname{diag}(\hat{X}) & UF(\hat{X})+F(\hat{X})U^T-e_ke_k^T\end{pmatrix}\begin{pmatrix}-2t & 0\\ -2tU^{-T}e & U^{-T}\end{pmatrix}\\
=&\begin{pmatrix}-2t\hat{\xi}+te^TU^{-1}\operatorname{diag}(\hat{X}) & t\operatorname{diag}(\hat{X})^T-2te^T(F(\hat{X})+U^{-1}F(\hat{X})U^T-U^{-1}e_ke_k^T)\\ tU^{-1}\operatorname{diag}(\hat{X}) & F(\hat{X})+U^{-1}F(\hat{X})U^T-U^{-1}e_ke_k^T\end{pmatrix}\begin{pmatrix}-2t & 0\\ -2tU^{-T}e & U^{-T}\end{pmatrix}\\
=&\begin{pmatrix}0 & 0\\ tU^{-1}\operatorname{diag}(\hat{X}) & F(\hat{X})+U^{-1}F(\hat{X})U^T-U^{-1}e_ke_k^T\end{pmatrix}\begin{pmatrix}-2t & 0\\ -2tU^{-T}e & U^{-T}\end{pmatrix}\\
=&\begin{pmatrix}0 & 0\\ 0 & F(\hat{X})U^{-T}+U^{-1}F(\hat{X})-U^{-1}e_ke_k^TU^{-T}\end{pmatrix}
\end{aligned} \tag{169}$$

with tridiagonal matrix

$$\begin{aligned}
&F(\hat{X})U^{-T}+U^{-1}F(\hat{X})-U^{-1}e_ke_k^TU^{-T}\\
=&\begin{pmatrix}\ddots & \ddots & & \\ \ddots & (2t-1)^{2(k-1-j)}(2(2t-1)(2t-\frac{2j-1}{2tk})+\frac{2j}{k}-4t^2) & (2t-1)^{2(k-1-i)}\frac{i}{k}(\frac{1}{2t-1}-\frac{1}{t}) & \\ & (2t-1)^{2(k-1-j)}\frac{j}{k}(\frac{1}{2t-1}-\frac{1}{t}) & \ddots & \ddots \\ & & \ddots & 1-\frac{1}{t(2t-1)}\frac{k-1}{k}\end{pmatrix}
\end{aligned} \tag{170}$$

If we diagonally scale the latter matrix in (170) with

$$D:=\operatorname{Diag}(((2t-1)^{-(k-1-j)})_{j=1,..,k}) \tag{171}$$

we see that the resulting tridiagonal matrix

$$\begin{aligned}
M:=&D(F(\hat{X})U^{-T}+U^{-1}F(\hat{X})-U^{-1}e_ke_k^TU^{-T})D\\
=&\begin{pmatrix}\ddots & \ddots & & \\ \ddots & (2(2t-1)(2t-\frac{2j-1}{2tk})+\frac{2j}{k}-4t^2) & \frac{i}{k}(\frac{1-t}{t}) & \\ & \frac{j}{k}(\frac{(1-t)}{t}) & \ddots & \ddots \\ & & \ddots & (2t-1)^2-\frac{2t-1}{t}\frac{k-1}{k}\end{pmatrix}
\end{aligned} \tag{172}$$

is weakly diagonally dominant and therefore positive semidefinite: With the convention that $M_{(-1)1}=\frac{1-1}{k}(\frac{(1-t)}{t})=0$ we can handle the first $k-1$ columns: Note that for $j\in\{1,..,k-1\}$

$$\begin{aligned}
&M_{jj}\geq|M_{(j-1)j}|+|M_{(j+1)j}|\\
\Leftrightarrow&(2(2t-1)(2t-\tfrac{2j-1}{2tk})+\tfrac{2j}{k}-4t^2)\geq\tfrac{j-1}{k}(\tfrac{(1-t)}{t})+\tfrac{j}{k}(\tfrac{(1-t)}{t})\\
\Leftrightarrow&4t^2-4t-\tfrac{4j-2}{k}+\tfrac{2j-1}{tk}+\tfrac{2j}{k}\geq\tfrac{2j-1}{k}(\tfrac{(1-t)}{t})\\
\Leftrightarrow&-4t(1-t)+\tfrac{1}{k}\geq 0
\end{aligned} \tag{173}$$

where the last inequalities hold true in this second case. For the last (k-th) column

$$\begin{aligned} & M_{k,k} \geq |M_{k-1,k}| \\ \Leftrightarrow & (2t-1)^2 - \tfrac{2t-1}{t}\tfrac{k-1}{k} \geq \tfrac{1-t}{t}\tfrac{k-1}{k} \\ \Leftrightarrow & (2t-1)^2 \geq \tfrac{k-1}{k} \end{aligned} \tag{174}$$

again the last inequality holds true for $1 \geq t > \frac{1}{2}(1+\sqrt{1-\frac{1}{k}})$ which concludes the second case and therefore our proof. ■

As we shall see next, similar proof techniques can be applied in a much more general setting.

4.2 Worst-Case-Complexity of Fixed Step Methods

The purpose of this subsection is to deal with the worst-case complexity of generalized iterations of the form (175), below, which we will refer to as fixed step methods (FSM). Let $T : \mathbb{H} \to \mathbb{H}$ be some nonexpansive operator and assume that $Fix(T) \neq \emptyset$. Define the residual mapping $g(x) := \frac{1}{2}(x - T(x))$. Let $x_* \in Fix(T)$, $x_1 \in \mathbb{H}$ and consider the iteration

$$x_i := x_1 + \sum_{j=1}^{i-1} t_{ij}(T(x_j) - x_j) = \quad x_1 - 2\sum_{j=1}^{i-1} t_{ij} g(x_j) \tag{175}$$

for fixed $k \in \mathbb{N}$ and $t_{ij} \in \mathbb{R}\ \forall i,j \in \{1,..,k\}$ with $i < j$. The special case $t_{ij} \equiv t$ corresponds to the KM iteration. Here we will consider the general case and show that the worst case complexity can always be derived from the solution of an SDP similar to the form of (152). In fact the only difference is the definition of the strict upper triangular matrix

$$L_t := -2 \begin{pmatrix} 0 & t_{12} & t_{13} & \dots & t_{1k} \\ 0 & 0 & t_{23} & \dots & t_{2k} \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & t_{k-1k} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{k \times k}.$$

and therefore $U_t := I - L_t$. The trick again relies on expressing the inequalities

$$\|g_i - g_j\|^2 \leq \langle g_i - g_j, x_i - x_j \rangle \tag{176}$$

and

$$\|g_i\|^2 \leq \langle g_i, x_i - x_* \rangle \tag{177}$$

in terms of the Gramian matrix $\begin{pmatrix} R^2 & b^T \\ b & A \end{pmatrix} \in \mathbb{S}_+^{k+1}$ for $R := \|\bar{x}_1 - x_*\| \geq 0$, $b := (\langle g_i, \bar{x}_1 - x_* \rangle)_{i=1}^k$ and $A := (\langle g_i, g_j \rangle)_{i,j=1}^k$ with $g_i := g(x_i)$. Now with the same reasoning as in the proof for the KM-iterations we can rewrite these inequalities as

$$\operatorname{diag}(AU_t)e^T + e\ \operatorname{diag}(U_t^T A)^T \leq AU_t + U_t^T A \tag{178}$$

respectively

$$\operatorname{diag}(AU_t) \leq b \tag{179}$$

and then obtain

$$\|g(x_k)\|^2 \leq \underset{y_0\in\mathbb{R}, y_1\in\mathbb{R}^k, Y_2=Y_2^T\in\mathbb{S}^k}{\text{maximize}} (Y_2)_{kk} \mid \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}_+^{k+1},\ y_0 \leq R^2,\ \operatorname{diag}(Y_2U_t) \leq y_1$$
$$\mid \operatorname{diag}(Y_2U_t)e^T + e\ \operatorname{diag}(U_t^TY_2)^T \leq Y_2U_t + U_t^TY_2 \tag{180}$$

via an SDP relaxation. It turns out that an optimal solutions of (180) gives us in fact something, that we will refer to as "worst-case-complexity" in the following sense: For the KM iteration we gave an example in order to show that our bound on the convergence rate is tight, i.e. could not be improved without making further assumptions. To enhance this reasoning we note that any feasible point of (180) actually corresponds to at least one iteration of the form (FSM) for some nonexpansive operator $\tilde{T}$ respectively one firmly nonexpansive operator $\tilde{g} := \frac{1}{2}(I - \tilde{T})$. Although not entirely accurate, one may think of a surjective function from the set of nonexpansive operators with fixed points to the feasible set of (180). As a consequence an optimal solution of (180) will in fact yield a bound that is not improvable in the sense that the inequality for $||g(x_k)||^2$ in (180) will be tight for appropriately chosen T and x_1. Now we prove our claim: Let $Y := \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix}$ be a feasible point of the SDP in (180) with $rank(Y) = d$. Then Y can be rewritten as $Y = \begin{pmatrix} \tilde{x}_1^T \\ G^T \end{pmatrix} \begin{pmatrix} \tilde{x}_1 & G \end{pmatrix}$ with $G = [\tilde{g}_1, ..., \tilde{g}_k] \in \mathbb{R}^{d\times k}$ and $\tilde{x}_1, \tilde{g}_1, ..., \tilde{g}_k \in \mathbb{R}^d$. If we now define $\tilde{x}_* := \tilde{g}_* := 0 \in \mathbb{R}^d$ and $\tilde{x}_i := \tilde{x}_1 - GL_te_i\ \forall i \in \{2, .., k\}$, then by construction and since Y is feasible for (180) the points $\tilde{x}_i, \tilde{g}_i$ will satisfy the equations $||\tilde{g}_i - \tilde{g}_j||_2^2 \leq (\tilde{g}_i - \tilde{g}_j)^T(\tilde{x}_i - \tilde{x}_j)$ for all $i, j \in I := \{*, 1, .., k\}$. The following Proposition tells us that we can find an operator extension to the whole space, i.e. a firmly nonexpansive operator $\tilde{g} : R^d \to \mathbb{R}^d$ with $\tilde{g}(\tilde{x}_i) = \tilde{g}_i\ \forall i \in I$.

Proposition 4.11. *Let $d, k \in \mathbb{N}_{\neq 0}$ be two positive integers. Define the set $I := \{*, 1, .., k\}$ with $k + 1$ elements. Let $\{x_i, g_i\}_{i\in I} \subset \mathbb{R}^d$ satisfy*

$$||g_i - g_j||_2^2 \leq (g_i - g_j)^T(x_i - x_j) \ \ \textit{for all}\ \ i, j \in I.$$

Then there exists a firmly nonexpansive operator $g : \mathbb{R}^d \to \mathbb{R}^d$, i.e.

$$||g(x) - g(y)||_2^2 \leq (g(x) - g(y))^T(x - y) \ \ \textit{for all}\ \ x, y \in \mathbb{R}^d. \tag{181}$$

that satisfies

$$g(x_i) = g_i\ \forall i \in I. \tag{182}$$

Proof. Note that by assumption $x_i = x_j$ implies $g_i = g_j$. This allows us to define the operator $\tilde{T} : \{x_*, x_1, ..., x_k\} \to \mathbb{R}^d$ via $\tilde{T}(x_i) := x_i - 2g_i \ \forall i \in I$. Also note, that $\tilde{T}$ is a nonexpansive mapping on its domain:

$$\begin{aligned} &||\tilde{T}(x_i) - \tilde{T}(x_j)||_2^2 - ||x_i - x_j||_2^2 \\ =&||x_i - x_j - 2(g_i - g_j)||_2^2 - ||x_i - x_j||_2^2 \\ =&4||g_i - g_j||_2^2 - 4(g_i - g_j)^T(x_i - x_j) \leq 0 \end{aligned} \tag{183}$$

holds true. By employing the Kirszbraun theorem (see [41] page 94 Hauptsatz I) we can extend $\tilde{T}$ to the whole space $\mathbb{R}^d$ in the following way: There exists a nonexpansive mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T(x_i) = \tilde{T}(x_i) \ \forall i \in I$. By our reasoning from before, we know that we can define the firmly nonexpansive operator $g : \mathbb{R}^d \to \mathbb{R}^d$ via $g(x) := \frac{1}{2}(x - T(x))$. Now g obviously satisfies

$$g(x_i) = \frac{1}{2}(x_i - T(x_i)) = \frac{1}{2}(x_i - \tilde{T}(x_i)) = \frac{1}{2}(x_i - (x_i - 2g_i)) = g_i \ \forall i \in I \tag{184}$$

which concludes our proof and also the argument from before. ∎

Remark 4.12. *Above we have proven that an optimal SDP solution of* (180) *yields an, in a certain sense, optimal or unimproveable bound, i.e. a worst-case complexity. However in a different sense these bounds may be very well improvable in at least three ways: First, additional knowledge about the structure of the operator T (for example if T is the derivative of some function), second, additional knowledge about the Hilbert space $\mathbb{H}$ (such as finite dimensionality) and third, asymptotically (for example the squared norm of residuals when using the KM iteration converges asymptotically with order $o(k^{-1})$ for $k \to \infty$ whenever the iterates converge) .*

Application 4.13. *Let us mention one obvious extension of our previous analysis. Let $F : \mathbb{H} \to \mathbb{H}$ be some L-cocoercive operator, i.e. there exists $L > 0$ such that*

$$\tfrac{1}{L} \|F(x) - F(y)\|^2 \leq \langle F(x) - F(y), x - y \rangle \tag{185}$$

is satisfied for all $x, y \in \mathbb{H}$. Assume that we are interested in finding a zero of F, i.e. $x_* \in \mathbb{H}$ such that $F(x_*) = 0$. Note that $\frac{1}{L}F(x)$ is firmly nonexpansive and therefore finding zeros of F is equivalent to finding fixed points of the nonexpansive operator $T(x) := x - \frac{2}{L}F(x)$. If we assume that there exists a zero of F, then we can estimate the worst case complexity of any FSM (175) *to find a zero of F. Specifically the squared norm $\|F(x_k)\|^2$ can be upper bounded by exploiting that the residual mapping $g(x) := \frac{1}{2}(x - T(x))$ is upper bounded by the SDP relaxation* (180) *and we have the identity $\|g(x)\|^2 = \left\|\frac{1}{2}(x - T(x))\right\|^2 = \frac{1}{L^2}\|F(x)\|^2$.*

4.2.1 Optimizing the Worst-Case-Complexity

As optimizers we are of course very much interested in the minimal worst-case complexity. If we see the SDP (180) not only as an optimization problem but also as a function of FSM step lengths, i.e. $\phi : \mathbb{R}^{\frac{k^2-k}{2}} \to \bar{\mathbb{R}}$ with

$$\begin{aligned}\phi(t) := \sup_{y_0\in\mathbb{R}, y_1\in\mathbb{R}^k, Y_2=Y_2^T\in\mathbb{S}^k} (Y_2)_{kk} \mid \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}_+^{k+1},\ y_0 \le R^2,\ \operatorname{diag}(Y_2U_t) \le y_1 \\ \mid \operatorname{diag}(Y_2U_t)e^T + e\ \operatorname{diag}(U_t^TY_2)^T \le Y_2U_t + U_t^TY_2\end{aligned} \tag{186}$$

minimizing the worst case becomes achievable by solving the optimization problem

$$\operatorname*{minimize}_{t\in\mathbb{R}^{\frac{1}{2}(k^2-k)}} \quad \phi(t) \tag{187}$$

globally. In fact for small values of k the function ϕ seems to be well behaved, i.e. continuous and possibly even (locally) Lipschitz-continuous (although not differentiable everywhere, as one can see from our proof of Theorem 4.9 for the KM-iteration). Now by using the local non-linear-programming (NLP) solvers min_f [49] respectively min_fc [31] (which are public domain and available under [50]) to minimize ϕ, together with YALMIP [54] and SeDuMi [78] to actually evaluate ϕ, for different values of k and random starting points, the final iterates often ended up "close" to the following FSM step lengths defined in (175)

$$t_{i,j}^{Halpern} := \tfrac{i}{j} \text{ for } i < j \tag{188}$$

giving raise to the so-called Halpern iteration [27] (which I showed in [53]): Starting with $x_1 = x_1^H$ and setting

$$x_{k+1}^H := \tfrac{1}{k+1}x_1 + (1 - \tfrac{1}{k+1})T(x_k^H). \tag{189}$$

In the paper [53], I used the dual problem (with $t = t^{Halpern}$) of (180) to show the following result:

Theorem 4.14. *Let $x_1 = x_1^H \in \mathbb{H}$ be arbitrary but fixed. If T has fixed points, i.e. $Fix(T) \neq \emptyset$, then the iterates defined in* (189) *satisfy*

$$\tfrac{1}{2}\left\|x_k^H - T(x_k^H)\right\| \le \frac{\|x_1 - x_*\|}{k} \quad \forall k \in \mathbb{N}_{\neq 0}, \forall x_* \in Fix(T) \tag{190}$$

It is not shown that this bound is optimal (although it is conjectured based on numerical evidence, i.e. on approximate solutions of (180) for $t = t^{Halpern}$). Based on numerical experiments with the local NLP-solvers, I believe that the Halpern step lengths (188) might actually yield a local (maybe even global) minimizer of (187). In practice however one often finds the convergence rate of the Halpern iteration inferior to the convergence rate of KM iteration. One reason for this phenomenon lies of course in the local or asymptotic behavior of both iteration: Note that when $T(x_k^H)$ is "close" to a fixed point x_* then the term $\frac{1}{k+1}x_1$ in the definition of the Halpern iteration will "pull" the iterates x_k away from x_*. Therefore the Halpern convergence rate will typically not be much faster then the upper bound predicted in above theorem. On the other hand the worst-case-complexity of the KM iteration (for $\|g(x_k)\|$) is much worse ($O(\sqrt{k^{-1}})$ instead of $O(k^{-1})$), but it *can* sometimes perform much better than our (unimprovable) bound in theorem 4.9 suggests in (146). Now if $t^{Halpern}$ indeed happens to be a global minimizer of ϕ, and if it is also unique, then the implication would be an unavoidable trade off between a "good" worst-case complexity and "fast" local/asymptotic convergence of FSM methods. In other words we could not find an FSM such that it will obtain an optimal worst-case complexity (in the sense of globally minimizing ϕ) and obtain fast local convergence at the same time.

4.2.2 Statistical Point of Entry

One of the main weak points of discussion about worst-case complexities is that their worst-case scenarios seem to occur rarely in practice. So the natural questions to ask is, whether we could not only quantify their likelihood, but also find an FSM that performs better in practice with high probability and maybe find a method that is optimized for an "average-case". In this section we will make a new, but small step in that direction. First note that the feasible set of (180) can be written as the union

$$\mathcal{F}^t := \bigcup_{0 \le r \le R} \mathcal{F}^t_r \tag{191}$$

of sliced sets

$$\mathcal{F}^t_r := \left\{ \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}^{k+1}_+ \mid_{\operatorname{diag}(Y_2U_t)e^T + e \ \operatorname{diag}(U_t^TY_2)^T \le Y_2U_t + U_t^TY_2}^{y_0 = r^2, \ \operatorname{diag}(Y_2U_t) \le y_1,} \right\} \tag{192}$$

which obviously satisfy $\mathcal{F}^t_0 = \{0\}$ and $\mathcal{F}^t_r = r^2\mathcal{F}^t_1$. As we have seen in Proposition 4.11 and the discussion before, each point in $\mathcal{F}^t$ corresponds to at least one possible iteration. We can therefore argue that the feasible set $\mathcal{F}^t$ holds (to a certain degree) not only information about the worst-case-complexity, but also about its likelihood and therefore about random-case- and average-case-complexities. If we sample a nonexpansive operator T with $Fix(T) \neq \emptyset$ at random: what can we expect how an FSM will behave? Allow us to be even sloppier here and rephrase the idea in the following way: Let $C \in \mathbb{S}^{k+1}$ be some a priori fixed matrix. If Ω is some set of possible outcomes and $Y : \Omega \to F^t_1$ is a random matrix obeying some probability distribution D. What is the probability distribution of $C \bullet Y$?

$$Y \sim D \Longrightarrow C \bullet Y \sim ???$$

One way to get an idea about the question above, is to employ Monte-Carlo Simulations: one can sample (pseudo-)random points $Y^{(1)}, ..., Y^{(p)}$ according to the distribution D from the set $\mathcal{F}^t_1$ and then check (and save) for each point $Y^{(l)}$ the quantity $C \bullet Y^{(l)}$ for $l \in \{1, .., p\}$. The main difficulty is of course sampling from $\mathcal{F}^t_1$. We recall the following well known (but improvable) possibility to generate samples obeying a uniform distribution on $\mathcal{F}^t_1$, namely a rejection strategy, to generate p samples in $\mathcal{F}^t_1$: First find a box $Q^t \subset \mathbb{S}^{k+1}$ with $\mathcal{F}^t_1 \subset Q$, i.e. $l^t_b, u^t_b \in \mathbb{R}^k, L^t_b, U^t_b \in \mathbb{S}^k$ such that

$$\mathcal{F}^t_1 \subset Q^t := \left\{ \begin{pmatrix} 1 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}^{k+1} \mid l^t_b \le y_1 \le u^t_b, L^t_b \le Y_1 \le U^t_b \right\} \tag{193}$$

the feasible set is fully contained inside the box. Now sample uniform distributed (pseudo-) random points from the box Q^t (which is simple). Check for each sample point $\tilde{Y} \in Q^t$ if

$\tilde{Y} \in \mathcal{F}_1^t$ is satisfied (not a strictly simple, but manageable task): if not, reject $\tilde{Y}$ and repeat; if yes, accept $\tilde{Y}$, save $C \bullet \tilde{Y}$ and stop if p samples have been accepted, otherwise repeat. For $C := \begin{pmatrix} 0 & 0^T \\ 0 & e_k e_k^T \end{pmatrix}$, i.e. the objective from (180) ($C \bullet \begin{pmatrix} 1 & y_1^T \\ y_1 & Y_2 \end{pmatrix} = (Y_2)_{kk}$), choosing $k = 2, p = 100000, t := t_{12} \in \{0.5, 0.6, .., 1\}$ the following MATLAB-plots of the empirical probability density function (PDF) and empirical cumulative density function (CDF) below were obtained.

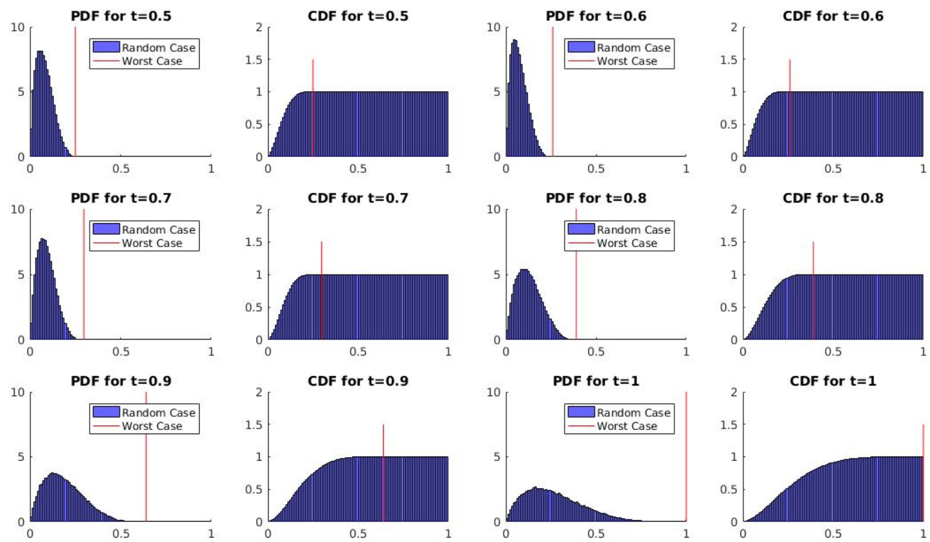


Figure 5: Empirical PDF and CDF plots for $(Y_2)_{22}$ ($\cong ||g(x_k)||^2$) for fixed $k = 2$ and different values of $t = t_{12}$. The red line marks the theoretical worst-case described earlier.

One possible interpretation of above plots is that the sliced feasible set F_1^t does not seem to contain "many" points that have an objective value close to the worst-case, i.e. the optimal solution of (180) seems to be a relatively "sharp" vertex of the feasible set. This does not allow any conclusions about the likelihood of worst-case scenarios (although we are certainly tempted to note, that they hint at a rather unlikely worst-case especially for larger t , we can not do that in good conscience). The main reason is that we did not consider a distribution on the feature space, i.e. the set of nonexpansive operators with fixed points, but only on F_1^t. In other words, we are not sure if or how the uniform distribution on F_1^t translates back to a meaningful probability distribution on the feature space. Another difficulty to overcome, lies within the proportion of F_1^t and Q_t: for $k = 2$ the sets F_1^t occupied less then 0.25% of the smallest boxes Q^t that I could find and for $k > 2$ this ratio drops even further (making the approach computationally expensive, because the ratio reflects the number of accepted samples $\tilde{Y}$). Although closer outer approximations, such as ellipsoids, could probably be used to ensure viability for slightly larger k, I am hoping in fact for computationally more attractive sampling strategies to arrive in the near future.

4.2.3 Selected Performance Criteria

So far we have focused on upper bounds on the squared norm of the residual of the last iterate ($\|g(x_k)\|^2$), but we can easily handle different performance criteria by changing the objective function of the SDP in (180). For example bounds on the quantity $\langle g(x_k), x_k - x_* \rangle$ can be realized by changing the objective to $e_k^T(y_1 + Y_2 L_t e_k)$. Some of these criteria are collected in the table below.

Criterion	Objective
$\|g(x_k)\|^2$	$(Y_2)_k$
$\sum_{i=1}^{k} \|g(x_i)\|^2$	$e^T \operatorname{diag}(Y_2)$
$\langle g(x_k), x_k - x_* \rangle$	$e_k^T(y_1 + Y_2 L_t e_k)$
$\|x_k - x_*\|^2$	$y_0 + e_k^T \operatorname{diag}(L_t^T(2y_1 e^T + Y_2 L_t))$

Table 6: Selected performance criteria and Objectives

We stretch the fact that some performance criteria will not trivially lead to new insights. For example if one is interested in bounds on the worst-case-performance of the squared distance between the last iterate x_k and the fixed point x_*, i.e. in the quantity $\|x_k - x_*\|^2$, then the worst-case will be attained with $\|x_k - x_*\|^2 = R^2 = \|x_1 - x_*\|^2$ and the corresponding nonexpansive operator is the identity mapping. However we can again try to employ the statistical point of view from the prior section this time for $C \bullet Y = y_0 + e_k^T \operatorname{diag}(L_t^T(2y_1 e^T + Y_2 L_t))$. As before I chose $k = 2, p = 100000, t := t_{12} \in \{0.5, 0.6, .., 1\}$ and ended up with the following MATLAB-plots of the empirical PDF and CDF below.

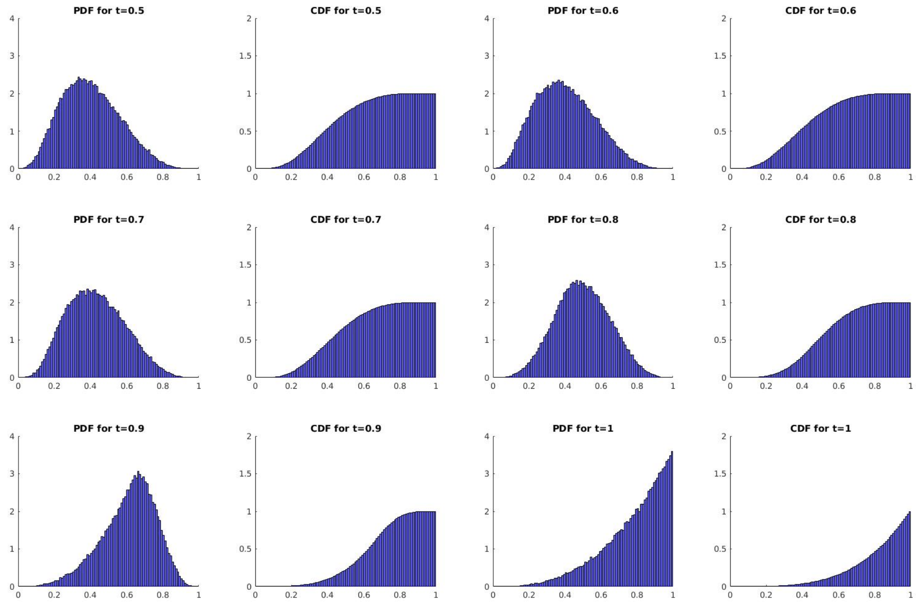


Figure 6: Empirical PDF and CDF plots for $y_0 + e_k^T \operatorname{diag}(L_t^T(2y_1e^T + Y_2L_t))$ for fixed $k = 2$ and different values of $t = t_{12}$. The theoretical worst-case is always equal to one.

Drawing conclusions from these plots is a challenge that we should rather leave others. We can not state with good conscience, that they hint at a decreasing worst-case probability as t increases towards values larger than $\frac{1}{2}$.
Another reason, for including the distance criterion into the table from before, is the extension of our worst-case-complexity from nonexpansive operators with fixed points to contracting operators in the next section.

4.3 Extension to Contractions

Although our main interest is finding fixed points of nonexpansive operators, our previous work can be very easily extended to q-contractions. In fact we can handle all Lipschitz-continuous operators with fixed points, but assuming contractions makes the existence assumption on the fixed point superfluous, which is why they are considered here. Let therefore $q \in [0, 1)$ and $T : \mathbb{H} \to \mathbb{H}$ be a q-contraction, i.e. satisfy

$$\|T(x) - T(y)\| \leq q \|x - y\| \tag{194}$$

for all $x, y \in \mathbb{H}$. As before we define the residual mapping $g(x) := \frac{1}{2}(x - T(x))$. If we rewrite (194) solely in terms of g, we obtain that g satisfies

$$\tfrac{1-q^2}{4} \|x - y\|^2 + \|g(x) - g(y)\|^2 \leq \langle g(x) - g(y), x - y\rangle \tag{195}$$

for all $x, y \in \mathbb{H}$. Let $x_1 \in \mathbb{H}$ be an arbitrary point and let $x_* \in \mathbb{H}$ be the unique fixed-point of T, respectively the unique zero of g (which exists by the Banach fixed point theorem).

Here we consider again an FSM

$$x_i := x_1 + \sum_{j=1}^{i-1} t_{ij}(T(x_j) - x_j) = \quad x_1 - 2\sum_{j=1}^{i-1} t_{ij} g(x_j) \tag{196}$$

for fixed $k \in \mathbb{N}$ and $t_{ij} \in \mathbb{R}\ \forall i, j \in \{1, .., k\}$ with $i < j$. Analogously to before we define $g_i := g(x_i) \forall i \in \{1, .., k\}$, as well as

$$L_t := -2 \begin{pmatrix} 0 & t_{12} & t_{13} & \dots & t_{1k} \\ 0 & 0 & t_{23} & \dots & t_{2k} \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & t_{k-1k} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{k \times k}.$$

and $U_t := I - L_t$. In order to express the inequalities in (195) we rewrite the quantities $\|x_i - x_j\|^2$ and $\|x_i - x_*\|^2$ in terms of the Gramian matrix $\begin{pmatrix} R^2 & b^T \\ b & A \end{pmatrix} \in \mathbb{S}_+^{k+1}$ for $R := \|\bar{x}_1 - x_*\| \geq 0$, the vector $b = (\langle g_i, \bar{x}_1 - x_* \rangle)_{i=1}^k$ and the matrix $A := (\langle g_i, g_j \rangle)_{i,j=1}^k$. Straightforward linear algebra calculations give us

$$\operatorname{diag}(L_t^T A L_t) e^T + e\ \operatorname{diag}(L_t^T A L_t)^T - 2 L_t^T A L_t = (\langle x_i - x_j, x_i - x_j \rangle)_{i,j=1}^k = (\|x_i - x_j\|^2)_{i,j=1}^k \tag{197}$$

and

$$\operatorname{diag}(R^2 e e^T + L_t^T b e^T + e b^T L_t + L_t^T A L_t) = \operatorname{diag}((\langle x_i - x_*, x_j - x_* \rangle)_{i,j=1}^k) = (\|x_i - x_*\|^2)_{i=1}^k \tag{198}$$

Let us define the linear and invertible operator $\mathcal{U}_t^q(A) := A - AL_t + \frac{1-q^2}{4} L_t^T A L_t$ (a generalization of our previous definition of multiplication with U_t for which we see that $\mathcal{U}_t^1(A) = AU_t$) is satisfied). We rewrite

$$\tfrac{1-q^2}{4} \|x_i - x_j\|^2 + \|g_i - g_j\|^2 \leq \langle g_i - g_j, x_i - x_j \rangle \tag{199}$$

by combining (178) and (197) as

$$\operatorname{diag}(\mathcal{U}_t^q(A)) e^T + e\ \operatorname{diag}(\mathcal{U}_t^q(A))^T \leq \mathcal{U}_t^q(A) + (\mathcal{U}_t^q(A))^T$$

and by using (179) and (198)

$$\tfrac{1-q^2}{4} \|x_i - x_*\|^2 + \|g_i\|^2 \leq \langle g_i, x_i - x_* \rangle \tag{200}$$

as

$$\operatorname{diag}(\mathcal{U}_t^q(A)) \leq b - \tfrac{(1-q^2)}{4}(R^2 e + 2 L_t^T b)$$

which we can again use to derive

$$\begin{aligned}
\|g(x_k)\|^2 = \underset{y_0\in\mathbb{R},y_1\in\mathbb{R}^k,Y_2=Y_2^T\in\mathbb{S}^k}{\text{maximize}} (Y_2)_{kk} \mid & \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}_+^{k+1},\ y_0 \le R^2 \\
& \mid \ \text{diag}(\mathcal{U}_t^q(Y_2)) \le y_1 - \tfrac{(1-q^2)}{4}(y_0 e + 2L_t^T y_1) \\
& \mid \ \text{diag}(\mathcal{U}_t^q(Y_2))e^T + e\ \ \text{diag}(\mathcal{U}_t^q(Y_2))^T \le \mathcal{U}_t^q(Y_2) + (\mathcal{U}_t^q(Y_2))^T \\
& \mid y_0 = R^2,\ y_1 = b,\ Y_2 = A \\
\le \underset{y_0\in\mathbb{R},y_1\in\mathbb{R}^k,Y_2=Y_2^T\in\mathbb{S}^k}{\text{maximize}} (Y_2)_{kk} \mid & \begin{pmatrix} y_0 & y_1^T \\ y_1 & Y_2 \end{pmatrix} \in \mathbb{S}_+^{k+1},\ y_0 \le R^2 \\
& \mid \ \text{diag}(\mathcal{U}_t^q(Y_2)) \le y_1 - \tfrac{(1-q^2)}{4}(y_0 e + 2L_t^T y_1) \\
& \mid \ \text{diag}(\mathcal{U}_t^q(Y_2))e^T + e\ \ \text{diag}(\mathcal{U}_t^q(Y_2))^T \le \mathcal{U}_t^q(Y_2) + (\mathcal{U}_t^q(Y_2))^T
\end{aligned} \tag{201}$$

where again the equality follows from construction and the inequality follows again from relaxation. Note that we could actually replace the constraint $y_0 \le R^2$ (where R^2 is a usually unknown constant) with for example $(Y_2)_{11} \le \|g(x_1)\|^2$ (where $\|g(x_1)\|^2$ is usually known) and still obtain a finite value for the SDP, in fact even the worst-case-complexity. With the same argument as in our discussion about nonexpansive operators, any feasible point with rank d of the SDP relaxation can be decomposed back into a set of $\{\tilde{x}_i, \tilde{g}_i\}_{i\in I} \subset \mathbb{R}^d$ for $I := \{*, 1, .., k\}$ satisfying (199) (with $\tilde{g}_i$ and $\tilde{x}_i$ instead of g_i and x_i). The following result ensures the existence of an operator extension to the full space $\mathbb{R}^d$.

Proposition 4.15. *Let $d, k \in \mathbb{N}_{\neq 0}$ be two positive integers. Define the set $I := \{*, 1, .., k\}$ with $k+1$ elements. Let $q \ge 0$ and let $\{x_i, g_i\}_{i\in I} \subset \mathbb{R}^d$ satisfy*

$$\tfrac{1-q^2}{4}||x_i - x_j||_2^2 + ||g_i - g_j||_2^2 \le (g_i - g_j)^T(x_i - x_j) \quad \textit{for all} \quad i, j \in I.$$

Then there exists an operator $g : \mathbb{R}^d \to \mathbb{R}^d$ that satisfies

$$\tfrac{1-q^2}{4}||x - y||_2^2 + ||g(x) - g(y)||_2^2 \le (g(x) - g(y))^T(x - y) \quad \textit{for all} \quad x, y \in \mathbb{R}^d \tag{202}$$

and

$$g(x_i) = g_i \ \forall i \in I. \tag{203}$$

Proof. Our proof here again uses the work of Kirszbraun [41]: Analogously to our prior proof we define the operator $\tilde{T} : \{x_*, x_1, ..., x_k\} \to \mathbb{R}^d$ via $\tilde{T}(x_i) := x_i - 2g_i \ \forall i \in I$. Note that this is well defined since $x_i = x_j$ implies $g_i = g_j$. Here the operator $\tilde{T}$ is a Lipschitz continuous mapping with constant q on its domain:

$$\begin{aligned}&||\tilde{T}(x_i)-\tilde{T}(x_j)||_2^2-q^2||x_i-x_j||_2^2\\=&||x_i-x_j-2(g_i-g_j)||_2^2-q^2||x_i-x_j||_2^2\\=&(1-q^2)||x_i-x_j||_2^2+4||g_i-g_j||_2^2-4(g_i-g_j)^T(x_i-x_j)\leq 0\end{aligned} \tag{204}$$

By employing the second Kirszbraun theorem (see [41] page 104 Hauptsatz 2 I): There exists a q-Lipschitz continuous mapping $T:\mathbb{R}^d\to\mathbb{R}^d$ such that $T(x_i)=\tilde{T}(x_i)\ \forall i\in I$. Now the operator $g:\mathbb{R}^d\to\mathbb{R}^d$ defined via $g(x):=\frac{1}{2}(x-T(x))$ will satisfy the claimed inequalities by reversing the calculations above. ■

Remark 4.16. *Proposition 4.11 and its generalization above were both stated for the space $\mathbb{R}^d$, but we like to emphasize that the results hold true for general (real) Hilbert spaces: Assuming the Hilbert space's inner product is in fact symmetric, the first proof part takes exactly the same form as above with $\mathbb{R}^d$ and its standard scalar product replaced. Finally the Lipschitz extension argument can be established by using Theorem 1.31. of [81].*

4.3.1 Zeros of Selected Strongly Monotone Operators

In this subsection we shift our focus again to finding zeros of operators $F:\mathbb{H}\to\mathbb{H}$. Let us start with stating and proving the following simple corollary of Proposition 4.15:

Corollary 4.17. *Let $d,k\in\mathbb{N}_{\neq 0}$ be two positive integers. Define the set $I:=\{*,1,..,k\}$ with $k+1$ elements. Let $0\leq\mu<L$ and let $\{x_i,F_i\}_{i\in I}\subset\mathbb{R}^d$ satisfy*

$$\frac{1}{1+\frac{\mu}{L}}(\mu||x_i-x_j||_2^2+\frac{1}{L}||F_i-F_j||_2^2)\leq(F_i-F_j)^T(x_i-x_j)\ \text{ for all }\ i,j\in I.$$

Then there exists an operator $F:\mathbb{R}^d\to\mathbb{R}^d$ with

$$F(x_i)=F_i\ \forall i\in I. \tag{205}$$

and such that $I-\frac{2}{L+\mu}F$ is a $\frac{L-\mu}{L+\mu}$-contraction (see (194)), and therefore (equivalently) F satisfies

$$\frac{1}{1+\frac{\mu}{L}}(\mu||x-y||_2^2+\frac{1}{L}||F(x)-F(y)||_2^2)\leq(F(x)-F(y))^T(x-y) \tag{206}$$

for all $x,y\in\mathbb{R}^d$.

Proof. Define $g_i:=\frac{1}{L+\mu}F_i$ and $q:=\frac{L-\mu}{L+\mu}$. Then we can rewrite the given inequalities as

$$\frac{1-q^2}{4}||x_i-x_j||_2^2+||g_i-g_j||_2^2\leq(g_i-g_j)^T(x_i-x_j)\ \text{ for all }\ i,j\in I.$$

By our previous Proposition 4.15 there exists a function $g:\mathbb{R}^d\to\mathbb{R}^d$ such that

$$\frac{1-q^2}{4}||x-y||_2^2+||g(x)-g(y)||_2^2\leq(g(x)-g(y))^T(x-y)\ \text{ for all }\ x,y\in\mathbb{R}^d \tag{207}$$

and

$$g(x_i)=g_i\ \forall i\in I. \tag{208}$$

If we now define $F(x) := (L+\mu)g(x)$ then $F(x_i) = (L+\mu)g(x_i) = (L+\mu)g_i = (L+\mu)\frac{1}{L+\mu}F_i = F_i$ and from $(g(x) = \frac{1}{L+\mu}F(x))$ we derive

$$\begin{aligned} 0 \geq & \frac{1-q^2}{4}||x-y||_2^2 + ||g(x)-g(y)||_2^2 - (g(x)-g(y))^T(x-y) \\ = & \underbrace{\frac{1-(\frac{L-\mu}{L+\mu})^2}{4}}_{=\frac{\mu L}{(L+\mu)^2}}||x-y||_2^2 + \frac{1}{(L+\mu)^2}||F(x)-F(y)||_2^2 - \frac{1}{(L+\mu)}(F(x)-F(y))^T(x-y) \end{aligned} \tag{209}$$

which yields the claimed inequality after a multiplication with $(L+\mu)$ and an appropriate rearrangement. ■

In [83] the inequalities from above corollary where identified as being necessary but not sufficient for L-smooth and μ-strongly convex interpolatability. Here we see that they are sufficient to yield an operator extension to the whole space. Let us recall that we say F is μ-strongly monotone if there exists $\mu \geq 0$ such that

$$\mu \|x-y\|^2 \leq \langle F(x) - F(y), x-y \rangle \tag{210}$$

is satisfied for all $x, y \in \mathbb{H}$. Also recall that we call F a gradient mapping, if there exists some function $f : \mathbb{H} \to \mathbb{R}$ with $F = \nabla f$. The original idea was to somehow "split" the contraction property into the concept of $L-$cocoercive (cf. (4.13)) and μ-strongly monotone operators. As it turns out, this contemplated "splitting" is not as straightforward as one might think: Consider for $0 \leq \mu < L$ the three statements:

(S0) F is a gradient mapping, L-Lipschitz and μ-strongly monotone

(S1) $I - \frac{2}{L+\mu}F$ is a $\frac{L-\mu}{L+\mu}$-contraction

(S2) F is μ-strongly monotone and L-cocoercive

Since we excluded the case $L = 0$, (S0) is basically the definition of an L-smooth and $\mu-$strongly convex function in the sense of [83], expressed solely in terms of its gradient. If F is a gradient mapping, then one can verify that (S1) and (S2) are equivalent (for the implication (S2) $\Rightarrow$ (S1) see for example Theorem 2.1.11 of [64]). Perhaps surprising is, that in general only the implication (S1) $\Rightarrow$ (S2) holds true (which we will show further down). Let us first, in order to show that the reverse implication does not hold in general, give the following example based on a Givens-Rotation:

Example 4.18. *Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined via $F(x) := Qx$ for*

$$Q := \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \tag{211}$$

Then for $|\theta| < \frac{\pi}{2}$ we calculate $(F(x)-F(y))^T(x-y) = (Q(x-y))^T(x-y) = \cos(\theta)||x-y||_2^2$ and see that F is $\cos(\theta)$-strongly monotone. Similar $||F(x)-F(y)||^2 = (x-y)^TQ^TQ(x-y) =$

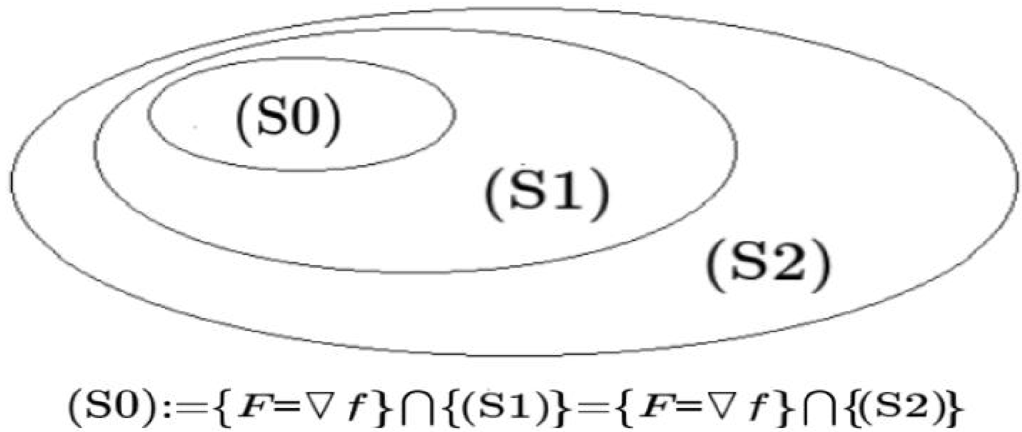


Figure 7: For gradient mappings conditions (S1) and (S2) are equivalent, although (S1) and (S2) are not equivalent in general.

$||x-y||^2$ *yields that* F *is* $\frac{1}{cos(\theta)}$*-cocoercive. In summary* F *satisfies (S2) with* $\mu := cos(\theta)$ *and* $L := \frac{1}{\cos(\theta)}$*, where both constants are chosen optimally. On the other hand, we obtain*

$$\begin{aligned}
&||x-y-\tfrac{2}{L+\mu}(F(x)-F(y))||^2-(\tfrac{L-\mu}{L+\mu})^2||x-y||^2\\
=&(x-y)^T(I-\tfrac{2}{L+\mu}(Q+Q^T)+(\tfrac{2}{L+\mu})^2Q^TQ)(x-y)-(\tfrac{L-\mu}{L+\mu})^2||x-y||^2\\
=&(1-\tfrac{4\mu}{L+\mu}+(\tfrac{2}{L+\mu})^2)||x-y||^2-(\tfrac{L-\mu}{L+\mu})^2||x-y||^2\\
=&4\tfrac{1-\mu^2}{(L+\mu)^2}||x-y||^2
\end{aligned} \tag{212}$$

and by setting $\theta := \arccos(\frac{1}{\sqrt{3}})$ *the factor* $4\frac{1-\mu^2}{(L+\mu)^2}$ *becomes equal to* $\frac{1}{2}$*, which implies that* F *does not satisfy (S1). (*$I-\frac{2}{L_F+\mu_F}F$ *is a* $\sqrt{4\frac{1-\mu^2}{(L+\mu)^2}+(\frac{L-\mu}{L+\mu})^2}=\sqrt{\frac{3}{4}}$*-contraction, but not a* $\sqrt{(\frac{L-\mu}{L+\mu})^2}=\frac{1}{2}$*-contraction.) Now on the other hand* F *is obviously, for* $L_F := \frac{1}{\cos(\theta)}+\sqrt{\frac{1}{\cos(\theta)^2}-1} > \frac{1}{\cos(\theta)} = L$ *and* $\mu_F := \frac{1}{\cos(\theta)}-\sqrt{\frac{1}{\cos(\theta)^2}-1} < \cos(\theta) = \mu$*, also* μ_F*-strongly monotone and* L_F*-cocoercive. Considering*

$$\begin{aligned}
&||x-y-\tfrac{2}{L_F+\mu_F}(F(x)-F(y))||^2-(\tfrac{L_F-\mu_F}{L_F+\mu_F})^2||x-y||^2\\
=&(x-y)^T(I-\tfrac{2}{L_F+\mu_F}(Q+Q^T)+(\tfrac{2}{L_F+\mu_F})^2Q^TQ)(x-y)-(\tfrac{L_F-\mu_F}{L_F+\mu_F})^2||x-y||^2\\
=&(1-\tfrac{4\cos(\theta)}{L_F+\mu_F}+(\tfrac{2}{L_F+\mu_F})^2)||x-y||^2-(\tfrac{L_F-\mu_F}{L_F+\mu_F})^2||x-y||^2\\
=&(1-\tfrac{8}{(L_F+\mu_F)^2}+(\tfrac{2}{L_F+\mu_F})^2)||x-y||^2-(\tfrac{L_F-\mu_F}{L_F+\mu_F})^2||x-y||^2\\
=&(1-(\tfrac{2}{L_F+\mu_F})^2-(\tfrac{L_F-\mu_F}{L_F+\mu_F})^2)||x-y||^2\\
=&\tfrac{4L_F\mu_F-4}{(L_F+\mu_F)^2}||x-y||^2\\
=&0
\end{aligned} \tag{213}$$

shows that $I - \frac{2}{L_F+\mu_F}F$ *is an* $\frac{L_F-\mu_F}{L_F+\mu_F} = |\sin(\theta)|$*-contraction which yields, for* $\theta = \arccos(\frac{1}{\sqrt{3}})$*, a significantly smaller (*$|sin(\arccos(\frac{1}{\sqrt{3}}))| = \sqrt{\frac{2}{3}}$ *) contraction constant. Also note that for any fixed* $\gamma \in \mathbb{R}$ $||(x-y) - \gamma(F(x) - F(y))||^2 \geq ||(x-y) - \frac{2}{L_F+\mu_F}(F(x) - F(y))||^2$ $\forall x, y \in \mathbb{R}^2$ *holds true, implying that the contraction constant can not be improved.*

Now let us show that statement (S1) implies (S2).

Lemma 4.19. *Let* $0 \leq \mu < L$*. If* $F : \mathbb{H} \to \mathbb{H}$ *satisfies (S1), i.e.* $I - \frac{2}{L+\mu}F$ *is a* $\frac{L-\mu}{L+\mu}$*-contraction, then it also satisfies (S2), i.e.* F *is* μ*-strongly monotone and* L*-cocoercive.*

Proof. Let us begin by proving μ-strong monotonicity of F: Analogously to the proof of Theorem 2.1.11 in [64], we can rewrite the inequality

$$\begin{aligned} 0 \geq & \left\| (x-y) - \tfrac{2}{L+\mu}(F(x) - F(y)) \right\|^2 - (\tfrac{L-\mu}{L+\mu})^2 \|x-y\|^2 \\ =& (1 - (\tfrac{L-\mu}{L+\mu})^2) \|x-y\|^2 - \tfrac{4}{L+\mu}\langle F(x) - F(y), x-y \rangle + \tfrac{4}{(L+\mu)^2} \|F(x) - F(y)\|^2 \end{aligned} \tag{214}$$

as

$$\tfrac{1}{L-\mu} \|F(x) - F(y) - \mu(x-y)\|^2 \leq \langle F(x) - F(y) - \mu(x-y), x-y \rangle \tag{215}$$

which implies $0 \leq \langle F(x) - F(y) - \mu(x-y), x-y \rangle = \langle F(x) - F(y), x-y \rangle - \mu \|x-y\|^2$, i.e. that F is μ-strongly monotone. By the Cauchy-Schwarz inequality (215) also implies

$$\|F(x) - F(y) - \mu(x-y)\| \leq (L-\mu) \|x-y\| \tag{216}$$

which we can, together with the triangle inequality, use

$$\begin{aligned} \|F(x) - F(y)\| =& \|F(x) - F(y) - \mu(x-y) + \mu(x-y)\| \\ \leq& \|F(x) - F(y) - \mu(x-y)\| + \mu \|x-y\| \\ \leq& (L-\mu) \|x-y\| + \mu \|x-y\| = L \|x-y\| \end{aligned} \tag{217}$$

to show that F is L-Lipschitz. By rewriting (214) once more and using the Lipschitz inequality as well as the Cauchy-Schwartz inequality, we conclude

$$\begin{aligned} \langle F(x) - F(y), x-y \rangle \geq & \tfrac{\mu L}{L+\mu} \|x-y\|^2 + \tfrac{1}{L+\mu} \|F(x) - F(y)\|^2 \\ \geq & \tfrac{\mu}{L+\mu} \|F(x) - F(y)\| \|x-y\| + \tfrac{1}{L+\mu} \|F(x) - F(y)\|^2 \\ \geq & \tfrac{\mu}{L+\mu} \langle F(x) - F(y), x-y \rangle + \tfrac{1}{L+\mu} \|F(x) - F(y)\|^2 \end{aligned} \tag{218}$$

and therefore

$$\langle F(x) - F(y), x-y \rangle \geq \tfrac{1}{L} \|F(x) - F(y)\|^2,$$

i.e. that F is $L-$cocoercive. ∎

As we have seen in our example, the reverse implication does not hold true in general. Now to make things even more complicated, note that while $I - \frac{2}{L+\mu}F$ might not be an $\frac{L-\mu}{L+\mu}$-contraction, it will still be some contraction with a different constant. Consider for example the following reasoning:

Lemma 4.20. *Let $0 < \mu < L$ and $F : \mathbb{H} \to \mathbb{H}$ be μ-strongly monotone and L-cocoercive. Then $I - \frac{2}{L+\mu}F$ is a $\sqrt{1-(\frac{2\mu}{L+\mu})^2}$-contraction.*

Proof.

$$\begin{aligned}
&\left\|(x-y) - \tfrac{2}{L+\mu}(F(x)-F(y))\right\|^2 \\
=& \|x-y\|^2 - \tfrac{4}{L+\mu}\langle F(x)-F(y), x-y\rangle + \tfrac{4}{(L+\mu)^2}\|F(x)-F(y)\|^2 \\
=& \|x-y\|^2 - \tfrac{4}{(L+\mu)^2}((\mu+L)\langle F(x)-F(y), x-y\rangle - \|F(x)-F(y)\|^2) \\
\le& \|x-y\|^2 - \tfrac{4}{(L+\mu)^2}(\mu\langle F(x)-F(y), x-y\rangle) \\
\le& \|x-y\|^2 - \tfrac{4}{(L+\mu)^2}(\mu^2\|x-y\|^2) \\
=& (1 - \tfrac{4\mu^2}{(L+\mu)^2})\|x-y\|^2
\end{aligned} \tag{219}$$

■

Note that the above contraction constant is attained in the first case of our example from before. In a way, this reflects two of the dilemmas that we are confronted with: Consider the following situation: Let us assume we are given some μ-strongly monotone and L-cocoercive operator $F : \mathbb{H} \to \mathbb{H}$ for which we want to find a zero, i.e. a point $x_* \in \mathbb{H}$ such that $F(x_*) = 0$ is satisfied. Obviously we can restate the problem of finding a zero, as finding a fixed point of $T_\gamma(x) := x - \gamma F(x)$. Our previous Lemma tells us that T_γ is actually a q-contraction for appropriate γ and q. We now could choose one of the many (possibly optimized) FSM, say $M(\gamma, q)$ in order to generate a sequence with $x_k \to x_*$. Let us explain the dilemma more carefully: assume our method $M(\gamma, q)$ is tuned for the given parameters γ and q. Who says that we have chosen them optimally? Our example clearly shows, that even knowing the maximal $0 < \mu$ and the minimal L, does not imply much knowledge about the optimal values of γ and q. Which brings us to our second dilemma: assume that μ and L are known to exist, but μ or L or both are not numerically available to us. Now what? Well there are a couple of standard ways to tackle these scenarios:

1. Educated guessing, although slightly unpopular, can sometimes just work. Guesses can come from application, rely on some additional structure or be simply empirical values.

2. Simulations (Trial and Error) may work, but suffer of course from additional computational effort and relies (more problematic) on a suitable simulation, which can be difficult to find.

3. Adaptive choice of parameters: If we define the optimal parameters

$$\mu^* := \inf_{x,y\in\mathbb{H},\ x\neq y} \frac{\langle F(x)-F(y), x-y\rangle}{\|x-y\|^2} \tag{220}$$

   and

$$\frac{1}{L^*} := \inf_{x,y\in\mathbb{H},\ F(x)\neq F(y)} \frac{\langle F(x)-F(y), x-y\rangle}{\|F(x)-F(y)\|^2} \tag{221}$$

it is easily seen that any two points $x, y \in \mathbb{H}$ with $F(x) \neq F(y)$ yield upper bounds on the optimal parameters. Two points $x, y \in \mathbb{H},\ x \neq y$ with $F(x) = F(y)$ imply that F can not be μ-strongly monotone for $\mu > 0$. Similarly we can consider the optimal contraction constant

$$q^* := \inf_{\gamma \in \mathbb{R}} \sup_{x,y \in \mathbb{H}, x \neq y} \frac{\|T_\gamma(x) - T_\gamma(y)\|}{\|x - y\|} \tag{222}$$

Note that from the Max-Min inequality we obtain

$$\begin{aligned} \inf_{\gamma \in \mathbb{R}} \sup_{x,y \in \mathbb{H}, x \neq y} \frac{\|T_\gamma(x) - T_\gamma(y)\|}{\|x - y\|} &\geq \sup_{x,y \in \mathbb{H}, x \neq y} \inf_{\gamma \in \mathbb{R}} \frac{\|T_\gamma(x) - T_\gamma(y)\|}{\|x - y\|} \\ &\geq \sup_{x,y \in \mathbb{H}, F(x) \neq F(y)} \sqrt{1 - (\tfrac{\langle F(x) - F(y), x - y \rangle}{\|F(x) - F(y\| \|x - y\|})^2} \end{aligned} \tag{223}$$

and again any two points $x, y \in \mathbb{H}$ with $F(x) \neq F(y)$ yield a lower bound on the optimal contraction constant. Here two points $x, y \in \mathbb{H},\ x \neq y$ satisfying $F(x) = F(y)$ results in $q^* = 1$.

In summary, answers to the questions of how to find worst case complexities for the class of L-cocoercive and μ-strongly monotone operators are difficult to obtain and may not be extremely satisfactory. This explains the word "selected" in the title of this section.

4.4 Extension to Complex Spaces

Being mathematicians, it is natural to ask what assumptions regarding the space $(\mathbb{H}, \langle ., . \rangle)$ are necessary for our analysis. For example we assumed the inner product to be symmetric, which basically limited everything to a real vector space. In principle conjugate symmetry ($\langle x, y \rangle = \overline{\langle y, x \rangle}$), i.e. complex Hilbert spaces, can be handled as well, but one has to take much greater care when it comes to inequalities. Let T be nonexpansive and define the residual mapping $g := \frac{1}{2}(I - T)$ as usual. We see that the original definition of a firmly nonexpansive operator makes only sense when $\langle g(x) - g(y), x - y \rangle$ is a real number. However, g does satisfy

$$\begin{aligned} 0 &\geq ||T(x) - T(y)||^2 - ||x - y||^2 \\ &= ||(x - y) - (T(x) - T(y)) - (x - y)||^2 - ||x - y||^2 \\ &= \langle (x - y) - (T(x) - T(y)) - (x - y), (x - y) - (T(x) - T(y)) - (x - y) \rangle - ||x - y||^2 \\ &= 4||g(x) - g(y)||^2 - 2\langle g(x) - g(y), x - y \rangle - 2\langle x - y, g(x) - g(y) \rangle \\ &= 4||g(x) - g(y)||^2 - 2\langle g(x) - g(y), x - y \rangle - \overline{2\langle g(x) - g(y), x - y \rangle} \end{aligned} \tag{224}$$

which we can interpret as a complex version of firm-nonexpansiveness. Also note that for a set of vectors $\{x_1 - x_*, g_1, ..., g_k\}$ an associated Gramian matrix will not be symmetric, but Hermitian positive semidefinite. As a consequence expressing an FSM in terms of that Gramian matrix, gets slightly more complicated, but is still possible. Note that in an SDP-Relaxation we have to consider the Hermitian complex matrices as a vector space over the real numbers (they are not a vector space over the complex numbers). Nevertheless one can check that each and every argument can be transfered, giving us the possibility to also consider worst-case complexities in complex Hilbert spaces.

What if we want to drop the assumption of working in a Hilbert space? Let us first note that Banach-spaces which satisfy the Parallelogram law, i.e.

$$\frac{1}{2}||x+y||^2 + \frac{1}{2}||x-y||^2 = ||x||^2 + ||y||^2 \tag{225}$$

do not pose a problem at all, because by using the Jordan von Neumann theorem, we can always define a scalar product that induces the norm and proceed analogously to the Hilbert space setting. On the other hand, Banach-spaces that do not satisfy the Parallelogram law, do pose a significant problem, as everything, we did requires the existence of a scalar product. Honestly I can not give a satisfactory answer at this point: For example note that the existence of an equivalent norm that is induced by a scalar product might still yield valid upper bounds but problematically since nonexpansiveness relies on the used norm, T might be a contraction in one norm but not in an equivalent one.

Leaving the extensions behind, let us conclude section four. In summary, the described approach for finding tight convergence rates in real or complex Hilbert spaces of any fixed step method is very attractive: We can use SDP solvers to find the optimal worst-case rates numerically and also (with a little bit of hard work) analytically. In addition we took the first step towards average-case optimized methods which may be taken further in the future.

5 Second Order Approach

Coming back to our original motivation of finding zeros of the gradient of the reduced Lagrangian, we might ask whether additional structure can help us achieving faster convergence. After all we have so far disregarded any curvature information of $\langle ., |.|_{\mathcal{K}}\rangle$ by setting $\mathcal{V}^{(k)}$, in our conceptual algorithm 2.24, equal to zero for all $k \in \mathbb{N}_0$. In this section we discuss the usage of (some) non-zero choices for $\mathcal{V}^{(k)}$. Now ideally we would find a rule to choose $\mathcal{V}^{(k)}$ that is well defined, ensures invertibility of $\mathcal{H} - \mathcal{V}^{(k)}$ for all $k \in \mathbb{N}_0$, is implementable and leads to faster (super-linear) convergence of 2.24 than the zero choice. While I do not think that this 'ideal' algorithm is achievable in general, it is achievable under suitable extra assumptions. It should be noted that in practice these methods are heuristic methods to some extent, as the extra assumptions may not be matched by reality. When dealing with second-order methods to find zeros of functions, we usually first think about Newton's method, i.e. (in $\mathbb{R}^n$) following (negative) directions that are a product of the inverse Jacobian times the function at the current iterate and repeating the procedure until a convergence criterion is satisfied. Recall that we are especially interested in finding zeros of

$$\nabla f(Z) = \tfrac{1}{2}\mathcal{H}[Z] - \tfrac{1}{2}|Z|_{\mathcal{K}} + R \tag{226}$$

and that its "Jacobian", i.e. the true second derivative (the "Hessian") of f does not necessarily exist.

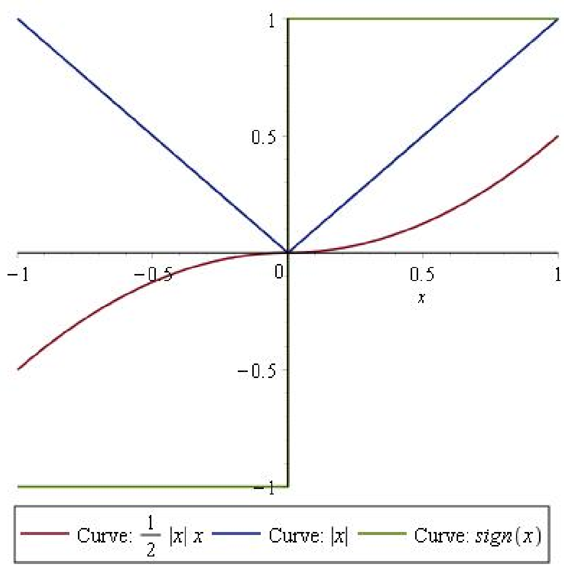


Figure 8: Maple Plot of the generalized absolute value, its antiderivative and its "derivative" for $\mathcal{K} = \mathbb{R}_+$. In zero there are multiple values shown, since the usual derivative is not well defined, the plot actually shows Clarke's generalized derivative.

The main idea is to now use a different but well defined operator instead of the possibly non-existing inverse second derivative.

5.1 Generalized Newton Approach

If we employ our previous notation of generalized derivatives, then a generalized Newton method could look like this:

$$Z^{(k+1)} := Z^{(k)} - \mathcal{M}_k^{-1}[\nabla f(Z^{(k)})] \tag{227}$$

where we choose the linear operator $\mathcal{M}_k \in \mathcal{L}(E, E)$ either from the Bouligand derivative $M_k \in \partial_B^2 f(Z^{(k)})$ or from Clarke's generalized derivative $M_k \in \partial_B^2 f(Z^{(k)})$. While the choosing rule is well defined, there is no reason for (227) to be well defined, because we have in general no control over the invertibility of elements from $\partial_B^2 f(Z^{(k)})$ or from $\partial^2 f(Z^{(k)})$. Recall that the sets have the forms

$$\partial_B^2 f(Z) = \tfrac{1}{2}\mathcal{H} - \tfrac{1}{2}\partial_B |Z|_{\mathcal{K}} \text{ and } \partial^2 f(Z) = \tfrac{1}{2}\mathcal{H} - \tfrac{1}{2}\partial |Z|_{\mathcal{K}}$$

and let us collect a few results on what we know about these generalized derivative sets. From Lipschitz continuity of a function it is easy to see that the Lipschitz constant gives us an upper bound on the operator norm of the derivative of that function whenever it exists. For us this especially implies the following bounds.

Lemma 5.1. *Let $Z \in E$ be arbitrary. Then*

$$||\mathcal{V}_B|| \leq 1 \text{ and } ||\tfrac{1}{2}(\mathcal{H} - \mathcal{V}_B)|| \leq 1 \;\forall\; \mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}} \tag{228}$$

as well as

$$||\mathcal{V}|| \leq 1 \text{ and } ||\tfrac{1}{2}(\mathcal{H} - \mathcal{V})|| \leq 1 \;\forall\; \mathcal{V} \in \partial |Z|_{\mathcal{K}} \tag{229}$$

hold true.

Proof. $|.|_{\mathcal{K}}$ is nonexpansive, i.e. 1-Lipschitz continuous which implies the first inequality. All other inequalities are then implied by the triangle inequality. ■

One might be wondering whether it makes a difference to choose a Bouligand or a Clarke derivative. While this might actually make a difference regarding condition numbers, it does not make any difference regarding invertibility, as the the following lemma shows.

Lemma 5.2. *Let $Z \in E$ be arbitrary. Then the elements $\frac{1}{2}(\mathcal{H} - \mathcal{V}_B) \in \partial_B^2 f(Z)$ are invertible for all $\mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}}$ if and only if $\frac{1}{2}(\mathcal{H} - \mathcal{V}) \in \partial^2 f(Z)$ are invertible for all $\mathcal{V} \in \partial |Z|_{\mathcal{K}}$.*

Proof. Because of the inclusion $\partial_B^2 f(Z) \subset \partial^2 f(Z)$ one direction is trivial. We prove the other direction by contradiction. Assume that all elements $\frac{1}{2}(\mathcal{H} - \mathcal{V}_B) \in \partial_B^2 f(Z)$ are regular, but $\frac{1}{2}(\mathcal{H} - \mathcal{V}) \in \partial^2 f(Z)$ is singular. Then there exists $\Delta Z \in E$ with $\|\Delta Z\| = 1$ and $\frac{1}{2}(\mathcal{H} - \mathcal{V})\Delta Z = 0$. By Caratheordorys theorem we can write $\mathcal{V} \in \partial |Z|_{\mathcal{K}}$ as $\mathcal{V} = \sum_{i=1}^{l} \lambda_i \mathcal{V}_i$ for some $\mathcal{V}_i \in \partial_B |Z|_{\mathcal{K}}$ and $\lambda_i > 0$ with $\sum_{i=1}^{l} \lambda_i = 1$. If we use compatibility of the operator norm well as our previous lemma, we see that

$$1 = \|\Delta Z\| = \|\mathcal{H}[\Delta Z]\| = \|\mathcal{V}[\Delta Z]\| = \left\| \sum_{i=1}^{l} \lambda_i \mathcal{V}_i[\Delta Z] \right\| \leq \sum_{i=1}^{l} \lambda_i \|\mathcal{V}_i[\Delta Z]\|$$

$$\leq \sum_{i=1}^{l} \lambda_i ||\mathcal{V}_i|| \, \|\Delta Z\| = \sum_{i=1}^{l} \lambda_i ||\mathcal{V}_i|| \leq \sum_{i=1}^{l} \lambda_i = 1$$

implies that in fact $\|\mathcal{V}_i[\Delta Z]\| = 1 \; \forall i \in \{1,..,l\}$ holds true. Therefore we can use the Cauchy Schwarz inequality

$$1 = \|\Delta Z\|^2 = \|\mathcal{H}[\Delta Z]\|^2 = \|\mathcal{V}[\Delta Z]\|^2 = \left\|\sum_{i=1}^{l} \lambda_i \mathcal{V}_i[\Delta Z]\right\|^2 = \sum_{i,j=1}^{l} \lambda_i\lambda_j \langle \mathcal{V}_i[\Delta Z], \mathcal{V}_j[\Delta Z]\rangle$$
$$\leq \sum_{i,j=1}^{l} \lambda_i\lambda_j \,\|\mathcal{V}_i[\Delta Z]\| \,\|\mathcal{V}_j[\Delta Z]\| = \sum_{i,j=1}^{l} \lambda_i\lambda_j = (\sum_{i=1}^{l} \lambda_i)^2 = 1$$

and obtain $\langle \mathcal{V}_i[\Delta Z], \mathcal{V}_j[\Delta Z]\rangle = 1 \forall i,j \in \{1,..,l\}$.

$$\left\|\tfrac{1}{2}(\mathcal{H} - \mathcal{V}_j)[\Delta Z]\right\| = \tfrac{1}{4}(\underbrace{\|\Delta Z\|}_{=1} - 2\langle \underbrace{\mathcal{H}[\Delta Z]}_{=\mathcal{V}[\Delta Z]}, \mathcal{V}_j[\Delta Z\rangle + \underbrace{\|\mathcal{V}_j[\Delta Z]\|}_{=1}$$
$$= \tfrac{1}{2}(1 - \underbrace{\langle \sum_{i=1}^{l} \lambda_i \mathcal{V}_i[\Delta Z], \mathcal{V}_j[\Delta Z]\rangle}_{=1} = 0$$

which yields a contradiction to the invertibility of $\frac{1}{2}(\mathcal{H} - \mathcal{V}_j)$. ■

The usual argument in order to show convergence of generalized Newton methods now starts with an assumption, namely that if all element of either $\partial_B^2 f(Z)$ or $\partial^2 f(Z)$ are invertible for some $Z \in E$, then there exists a neighborhood of $N_B \subset E$ of Z on which the generalized derivatives are also invertible. Let us adapt the notation of the according result from [71] Lemma 2.6., where this property is called strong BD-regularity.

Lemma 5.3. *Let $X \in E$ be arbitrary. If all elements of $\partial_B^2 f(X)$ are invertible, then there exists a neighborhood N_B of X and a constant $\gamma_B > 0$, such that for any $Y \in N_B$ and $\mathcal{M}_B \in \partial_B^2 f(Y)$, $\mathcal{M}_B$ is nonsingular and*

$$||\mathcal{M}_B^{-1}|| \leq \gamma_B \tag{230}$$

Together with our previous lemma we can state the following corollary specific to our situation.

Corollary 5.4. *Let $X \in E$ be arbitrary. If all elements of $\partial_B^2 f(X)$ are invertible, then there exists a neighborhood N of X and a constant $\gamma > 0$, such that for any $Y \in N$ and $\mathcal{M} \in \partial^2 f(Y)$, $\mathcal{M}$ is nonsingular and*

$$||\mathcal{M}^{-1}|| \leq \gamma \tag{231}$$

Proof. According to Lemma 5.3. there exists a neighborhood N_B of X and and constant γ_B such that for any $Y \in N_B$ and $\mathcal{M}_B \in \partial_B^2 f(Y)$, $\mathcal{M}_B$ is nonsingular and $||\mathcal{M}_B^{-1}|| \leq \gamma_B$. By using Lemma 5.2 we know that any $\mathcal{M} \in \partial^2 f(Y)$ must also be invertible. Therefore the inverse operator function $inv : \partial^2 f(Y) \to \mathcal{L}(E,E), \mathcal{M} \mapsto \mathcal{M}^{-1}$ is continuous. Since $\partial^2 f(Y)$ is compact this implies that $\gamma_Y := \sup_{\mathcal{M} \in \partial^2 f(Y)} ||\mathcal{M}^{-1}||$ is finite. We can now choose sufficiently small $\epsilon > 0$ such that the (closed) ϵ-ball $B_\epsilon(X)$ around X satisfies $B_\epsilon(X) \subset N_B$

and γ_Y is bounded on the intersection $N_B \cap B_\epsilon(X)$, i.e. $\gamma := \sup_{Y \in N_B \cap B_\epsilon(X)} \gamma_Y$ is finite. Otherwise there would exist a sequence $\{Y_k\}_{k\in\mathbb{N}}$ converging to X and a sequence $\{\mathcal{M}_k\}_{k\in\mathbb{N}}$ with $||\mathcal{M}_k^{-1}|| \to \infty$. Moreover $\mathcal{M}_k \in \partial^2 f(Y_k)$ is bounded in norm, and therefore without loss of generality convergent to some $\bar{\mathcal{M}}$. By using the definition of Clarke's generalized derivative we see that $\bar{\mathcal{M}} \in \partial^2 f(X)$ is satisfied, yielding a contradiction to the finiteness of γ_X. Specifically we obtain for all $Y \in N := int(N_B \cap B_\epsilon(Z))$ the claimed inequality. ■

Remark 5.5. *Above corollary is highly dependent on the special form of $\partial^2 f$ and does not hold in general, as one can easily see from the normal absolute value function. Also note that we can replace the open neighborhoods N_B and N with for example a closed ϵ-ball (for sufficiently small $\epsilon > 0$) and then relate the "smallest" constants from lemma 5.3 and Lemma 5.4 via $\gamma_B \leq \gamma$. Even on the same neighborhood there does not seem to be a reason for the "smallest" constants to be equal though, and I think that examples can be constructed such that they are not. This could potentially affect the size of any area of attraction in a local convergence result.*

Additionally to invertibility, we make another assumption regarding the approximation quality of the generalized derivatives in order to show local convergence of (227) (see [44] 2.3. for an example of non-convergence of the generalized Newton method). Specifically one assumption to be considered is the situation where $Z^{(*)} \in E$ satisfies $\nabla f(Z^{(*)}) = 0$ and for any $\mathcal{V} \in \partial|Z^{(*)} + \Delta Z|_\mathcal{K}$ the relation

$$|Z^{(*)} + \Delta Z|_\mathcal{K} - |Z^{(*)}|_\mathcal{K} - \mathcal{V}[\Delta Z] \in o(\|\Delta Z\|) \text{ for } \Delta Z \to 0 \tag{232}$$

holds. Relation (232) was discussed in a more general setting in [42] (theorem 2). More importantly though, it has also been identified to be essentially[12] a necessary condition for local convergence. Here we choose to use the slightly stronger, and more common, concept of semismoothness, which was introduced for functionals in [59] and later extended to vector-valued functions [70]. Although our functions here are not vector valued in the usual sense, we can still think of them as such by employing an isometric isomorphism. Therefore all known results regarding semi-smoothness can be applied in our situation. Let us translate this property for our particular case.

Definition 5.6. *For $Z \in E$ we say that $|.|_\mathcal{K}$ is semismooth at Z if the limit*

$$\lim_{\substack{\mathcal{V} \in \partial|Z + t\Delta Z'|_\mathcal{K} \\ \Delta Z' \to \Delta Z,\ t \downarrow 0}} \mathcal{V}[\Delta Z'] \tag{233}$$

exists for all $\Delta Z \in E$.

The general definition includes local Lipschitz continuity as a requirement, which is of course globally fulfilled by $|.|_\mathcal{K}$. In order to make above definition "usable", we need to introduce the following notation of directional derivatives

$$|Z|'_\mathcal{K}(\Delta Z) := \lim_{t \downarrow 0} \frac{|Z + t\Delta Z|_\mathcal{K} - |Z|_\mathcal{K}}{t} \tag{234}$$

[12] If one uses Clarke's generalized derivative in the generalized Newton method for finding a zero of a Lipschitzian homeomorphism. See theorem 3 of [42]

of $|.|_{\mathcal{K}}$ at $Z \in E$ in direction ΔZ. Note that above limit does not need to exist in general, but semismoothness actually implies its existence, as we see by considering the following well known equivalences (see e.g. [70], [72] or [29] for proofs of a general version).

Lemma 5.7. *The following statements are equivalent:*

1. $|.|_{\mathcal{K}}$ *is semismooth at* Z
2. $|Z|'_{\mathcal{K}}(\Delta Z)$ *exists for all* $\Delta Z \in E$ *and for any* $\mathcal{V} \in \partial |Z + \Delta Z|_{\mathcal{K}}$

$$\left\| \mathcal{V}[\Delta Z] - |Z|'_{\mathcal{K}}(\Delta Z) \right\| = o(\|\Delta Z\|) \tag{235}$$

   for $\Delta Z \to 0$

3. $|Z|'_{\mathcal{K}}(\Delta Z)$ *exists for all* $\Delta Z \in E$ *and for any* $\mathcal{V} \in \partial |Z + \Delta Z|_{\mathcal{K}}$

$$\| |Z + \Delta Z|_{\mathcal{K}} - |Z|_{\mathcal{K}} - \mathcal{V}[\Delta Z] \| = o(\|\Delta Z\|) \tag{236}$$

   for $\Delta Z \to 0$

The last characterization (for $Z = Z^{(*)}$) obviously implies (232), but it is strictly stronger because in general $|.|_{\mathcal{K}}$ does not need to be directional differentiable (for an example of a convex projection that is not directional differentiable everywhere, see for [76]). Another reasonable and even stronger assumption is the concept of γ-order semismoothness.

Definition 5.8. *For* $0 < \gamma \leq 1$ *the generalized absolute value* $|.|_{\mathcal{K}}$ *is called* γ*-order-semismooth at* $Z \in E$ *if* $|Z|'_{\mathcal{K}}(\Delta Z)$ *exists for all* $\Delta Z \in E$ *and for any* $\mathcal{V} \in \partial |Z + \Delta Z|_{\mathcal{K}}$

$$\| |Z + \Delta Z|_{\mathcal{K}} - |Z|_{\mathcal{K}} - \mathcal{V}[\Delta Z] \| = o(\|\Delta Z\|^{1+\gamma}) \tag{237}$$

for $\Delta Z \to 0$. *We call* $|.|_{\mathcal{K}}$ *strongly semismooth if it is* 1*-order semismooth.*

Now luckily neither semismoothness nor γ-order semismoothness is a very restrictive assumption in our case. For example the generalized absolute value is strongly semismooth if the cone $\mathcal{K}$ is either the nonnegative orthant (trivial exercise), the second-order cone ([12] proposition 4.3) or the semidefinite cone ([80] theorem 4.13). In fact a function is (γ-order) semismooth if and only if its components are so(see e.g. [89] proposition 3.6) and therefore the generalized absolute value of any direct product of the above cones will be strongly semismooth. Let us write out the generalized Newton iteration without the index k. If we assume that $Z^{(*)} \in E$ satisfies $\nabla f(Z^{(*)}) = 0$ and $\frac{1}{2}(\mathcal{H} - \mathcal{V}) =: \mathcal{M} \in \partial^2 f(Z)$ to be invertible we obtain

$$\begin{aligned} Z - (\mathcal{M})^{-1}[\nabla f(Z)] - Z^{(*)} &= \mathcal{M}^{-1}[\mathcal{M}[Z - Z^{(*)}] - \nabla f(Z)] \\ &= \mathcal{M}^{-1}[\tfrac{1}{2}(\mathcal{H} - \mathcal{V})[Z - Z^{(*)}] - (\nabla f(Z) - \nabla f(Z^{(*)}))] \\ &= \mathcal{M}^{-1}[\tfrac{1}{2}(\mathcal{H} - \mathcal{V})[Z - Z^{(*)}] - \tfrac{1}{2}(\mathcal{H}[Z - Z^{(*)}] - (|Z|_{\mathcal{K}} - |Z^{(*)}|_{\mathcal{K}})] \\ &= \tfrac{1}{2}\mathcal{M}^{-1}[|Z|_{\mathcal{K}} - |Z^{(*)}|_{\mathcal{K}} - \mathcal{V}[Z^{(*)} - Z]] \end{aligned} \tag{238}$$

and therefore

$$\left\| Z - (\mathcal{M})^{-1}[\nabla f(Z)] - Z^{(*)} \right\| \leq \tfrac{1}{2}||\mathcal{M}^{-1}|| \left\| |Z|_{\mathcal{K}} - |Z^{(*)}|_{\mathcal{K}} - \mathcal{V}[Z^{(*)} - Z] \right\| \tag{239}$$

holds true. If we assume that Z approaches $Z^{(*)}$, $Z \to Z^{(*)}$ and we further assume $\mathcal{M} \in \partial^2 f(Z)$ to remain bounded, the rate of convergence relies greatly on the level of semismoothness of $|.|_{\mathcal{K}}$ and in fact the following result is true.

Theorem 5.9. *Let $\nabla f : E \to E$ be defined as in* (226)*, suppose that $Z^{(*)} \in E$ satisfies $\nabla f(Z^{(*)}) = 0$ and that all elements of $\partial_B^2 f(Z^{(*)})$ are invertible. Then the generalized Newton method* (227) *converges Q-superlinearly in a neighborhood of $Z^{(*)}$ if $|.|_{\mathcal{K}}$ is semismooth in $Z^{(*)}$ and it converges quadratically if $|.|_{\mathcal{K}}$ is strongly semismooth in $Z^{(*)}$.*

Proof. According to lemma 5.2 all elements of $\partial^2 f(Z^{(*)})$ are invertible. Since (strong) semismoothness is closed under scalar multiplication as well as summation and $\mathcal{H}[Z]$ is continuously differentiable and therefore strongly semismooth (see e.g. [89] propositions 3.4 and 3.5), ∇f is semismooth at $Z^{(*)}$. The claim then follows from [71](theorem 3.1) if all operators $\mathcal{M}_k$ are chosen from $\partial_B^2 f(Z^{(k)})$ or from [70] (theorem 3.2) if they are choosen from $\partial^2 f(Z^{(k)})$. ■

Above theorem reflects the opportunities that the generalized Newton method yields as well as its limitations. The fast local convergence relies on a certain amount of regularity and on an approximation property. While semismoothness will in practice not really be an issue, the locality argument and regularity assumption might. Let us address a possible solution for the regularity issue first. One attempt to consider is of course the usage of pseudo-inverses. Our previous theory fits another, and apparently well studied, approach much better, which we briefly present here. Consider the regularized function

$$\nabla f_\lambda(Z) := (1-\lambda)\nabla f(Z) + \tfrac{\lambda}{2}\mathcal{H}[Z - Z^{(0)}] \tag{240}$$

for arbitrary but fixed $Z^{(0)} \in E$. It is easy to see that

$$\begin{aligned} T_\lambda(Z) :&= Z - 2\mathcal{H}[\nabla f_\lambda(Z)] \\ &= \lambda Z^{(0)} + (1-\lambda)\mathcal{H}[|Z|_{\mathcal{K}} - 2\mathcal{R}] \end{aligned} \tag{241}$$

is a contraction for every $\lambda \in (0,1]$. By the Banach contraction theorem we conclude that T_λ has a unique fixed point $Z_\lambda \in E$ and therefore ∇f_λ a unique zero for every $\lambda \in (0,1]$. We will refer to the curve $\{Z_\lambda\} \subset E$ as Halpern implicit iteration, following the recent phd thesis [58]. Note that the Halpern implicit iteration has been extensively studied by Browder (see for example [6], [7]) and it is especially known ([8]) that, if $Fix(T_0) \neq \emptyset$, than Z_λ converges to the fixed point closest to $Z^{(0)}$ for $\lambda \to 0$. Note that ∇f_λ is semismooth if and only if ∇f is too. The difference is that all elements of $\partial \nabla f_\lambda(Z) = \frac{1}{2}\mathcal{H} - (1-\lambda)\partial |Z|_{\mathcal{K}}$ are invertible for $\lambda \in (0,1]$, implying that we can in principle use the generalized Newton method to "follow" the curve for $\lambda \searrow 0$, even when some of the elements of $\partial^2 f(Z^{(*)}) = \partial \nabla f_0(Z^{(*)})$ are not invertible. In order to globalize the generalized Newton method, we can use a damped version, i.e. we modify the iteration (227)

$$Z^{(k+1)} := Z^{(k)} - t_k \mathcal{M}_k^{-1}[\nabla f(Z^{(k)})] \tag{242}$$

by introducing step lengths $t_k \in (0,1]$, which can for example be chosen with respect to the merit function $\frac{1}{2}\|\nabla f(Z)\|^2$. Similarly, for the implicit Halpern iteration we can consider

$$Z_\lambda^{(k+1)} := Z_\lambda^{(k)} - t_{\lambda k} \mathcal{M}_{\lambda k}^{-1}[\nabla f_\lambda(Z_\lambda^{(k)})] \tag{243}$$

with $M_{\lambda k} \in \partial\nabla f_\lambda(Z_\lambda^{(k)})$, $t_{\lambda_k} \in (0,1]$ and analogous merit function. Now the real issue of (generalized) Newton methods is the high computational cost of solving the linear system. First, we do know explicit representations of Clarke's generalized derivative in some cases (including the non-negative orthant, the second order cone (see [38] Lemma 2.6) and the semidefinite cone (see [55] or [48]), but in general finding and choosing elements from the generalized derivative of the generalized absolute value can be much harder than evaluating the generalized absolute value itself (as we will see in the next section). Second even if we obtain an explicit representation, the cost of actually solving systems of the form $\frac{1}{2}(\mathcal{H}-\mathcal{V})[\Delta Z] = rhs$ may be out of scale. For example in the semidefinite case ($\mathcal{K} = \mathbb{S}^n_+$) the linear operator $\frac{1}{2}(\mathcal{H}-\mathcal{V})$ can be regarded as a symmetric matrix in $\mathbb{S}^{\frac{n^2+n}{2}} \subset \mathbb{R}^{\frac{n^2+n}{2}\times\frac{n^2+n}{2}}$. Direct methods (such as LU- or LDLT-decompositions with backward substitution) would potentially need $O((\frac{n^2+n}{2})^3) = O(n^6)$ floating point operations (as well as lots of memory), which makes their application virtually impossible even for medium sized $n \in \mathbb{N}$. We could employ iterative methods such as Minres [68] or MINRES-QLP [13], but their convergence rate relies heavily on the condition numbers of the linear systems (which might be very large or even infinite). Let us conclude this section and state, that from a practical point of view, the generalized Newton method is promising as long as the linear systems can be solved quickly. Unfortunately we have no real control over their explicit forms nor their condition numbers, as they are extremely problem dependent. In the next section we derive a new technique that tries to avoid this complication and appears to be a reasonable compromise between the KM iteration and the generalized Newton method, based on quasi Newton methods.

5.2 Norm Constrained Quasi Newton Approach

Let us start by recalling the general idea of direct quasi Newton methods. Let us for simplicity assume that we are given a differentiable function $F : \mathbb{R}^n \to \mathbb{R}^n$. Assume further that we are interested in finding a zero $x_* \in \mathbb{R}^n$ of F, but we are only given an (arbitrary) point $x^k \in \mathbb{R}^n$. Unlike the normal Newton method we do not explicitly use the (inverse) derivative of F, but an (invertible) approximation $B_k \in \mathbb{R}^{n\times n}$ to the (averaged) Jacobian to perform one step. After taking this step

$$x^{k+1} := x^k - B_k^{-1}F(x_k)$$

we can update the approximation B_k by exploiting the secant equation. This is usually done by defining B_{k+1} as the solution of the optimization problem

$$\begin{aligned} \underset{\tilde{B}\in\mathbb{R}^{n\times n}}{\text{minimize}} \quad & ||M_k(\tilde{B} - B_k)M_k||_F \\ \text{subject to} \quad & \tilde{B}\tilde{s}_k = \tilde{g}_k \\ & (\tilde{B} = \tilde{B}^T) \end{aligned}$$

for some (invertible) weight matrix $M_k \in \mathbb{R}^{n\times n}$, $\tilde{g}_k := F(x_{k+1}) - F(x_k)$ and $\tilde{s}_k := x_{k+1} - x_k$, which can be solved explicitly[13]. One drawback however which is easily seen, is that this

[13]For example, if $M_k = M = M^T$ is symmetric invertible, $g = \tilde{g}_k$ and $s = \tilde{s}_k$ then the solution is given by $B_+ = B + \frac{(g-Bs)c^T + c(g-Bs)^T}{c^Ts} - \frac{(g-Bs)^Ts}{(c^Ts)^2}cc^T$ for $c = M^{-2}s$, see for example [33].

update procedure ignores lots of (possibly available) information about the true Jacobian of F. For example assume that we are trying to approximate a symmetric Jacobian and for some reason (e.g. due to some known Lipschitz constant) we know that eigenvalues of the true (averaged) Jacobian are contained in an interval $[\tilde{\mu}, \tilde{L}]$ for some $\tilde{\mu},\ \tilde{L} \in \mathbb{R}$ with $\tilde{\mu} \leq \tilde{L}$. In this case, we might be tempted to solve

$$\begin{aligned} \underset{\tilde{B}\in\mathbb{R}^{n\times n}}{\text{minimize}} \quad & ||M_k(\tilde{B} - B_k)M_k||_F \\ \text{subject to} \quad & \tilde{B}\tilde{s}_k = \tilde{g}_k \\ & \tilde{B} = \tilde{B}^T \\ & \tilde{\mu} I \preceq \tilde{B} \preceq \tilde{L} I \end{aligned} \tag{244}$$

instead, but unfortunately there does not seem to be an explicit formula for the solution and in general the application of a general purpose SDP solver tends to be unreasonably expensive. We will see shortly that we can dramatically decrease the computational cost if B_k in (244) is of (shifted) low rank. Note that B_k needs to be of full rank for the outer method to be well defined. In our case, the trick is to assume that the true Jacobian of F is a linear combination of a known (symmetric) Matrix $H(x)$ and another (unknown) term $V(x)$ (for example $\partial^2 f = \frac{1}{2}(H - \partial|.|_{\mathcal{K}})$). If our approximation B_k has inherited this structure, for example if

$$B_k = \tfrac{1}{2}(H - V_k) \tag{245}$$

then we only need a new approximation V_{k+1} of the unknown term and can define the new approximation of the Jacobian immediately (for example via $B_{k+1} := \frac{1}{2}(H - V_{k+1})$). Let us also assume (known) bounds on the eigenvalues of the unknown term ($\partial|.|_{\mathcal{K}}$), say $\mu, L \in \mathbb{R}$ with $\mu \leq L$. Instead of (244) we can now consider

$$\begin{aligned} \underset{\tilde{V}\in\mathbb{R}^{n\times n}}{\text{minimize}} \quad & ||M_k(\tilde{V} - V_k)M_k||_F \\ \text{subject to} \quad & \tilde{V}s_k = g_k \\ & \tilde{V} = \tilde{V}^T \\ & \mu I \preceq \tilde{V} \preceq L I \end{aligned} \tag{246}$$

for adapted s_k, g_k. The difference between (244) and (246) is that it may make sense to assume a low rank structure of V_k (but it does not really for B_k). Let us now show how we can exploit such a low rank of structure V_k. For a possible application in convex unconstrained optimization we will actually go a step further and include a shifted low rank structure (which will not be used here, but might be useful in other situations). If the known matrix $H(x)$ is a multiple of the identity, then one can skip (246) and immediately consider (247) below. I am uncertain, whether the result from the next lemma has been noted before or not. There exist so called self-scaling variable metric algorithms, see for example [66], but I am not aware that this has been exploited in any practical algorithm. We will, for now, drop the index k, assume that we are given constants $\mu \leq \gamma \leq\ L$ as well as vectors $s, y \in \mathbb{R}^n$, an invertible symmetric matrix $M = M^T \in \mathbb{R}^{n\times n}$ and consider

$$\begin{aligned} V_+ :=\underset{\tilde V\in\mathbb{R}^{n\times n}}{\text{minimize}} \quad & ||M(\tilde V - V_-)M||_F \\ \text{subject to} \quad & \tilde V s = g \\ & \tilde V = \tilde V^T \\ & \mu I \preceq \tilde V \preceq LI \end{aligned} \tag{247}$$

instead of (246).

Lemma 5.10. *Assume that* (247) *is feasible and that* $V_- = \gamma I + P_-^T \hat W_- P_-$*, for some* $P_- \in \mathbb{R}^{p\times n}$ *and* $\hat W_- = \hat W_-^T \in \mathbb{R}^{p\times p}$ *with* $\mu \leq \gamma \leq L$ *are given. Let us define* $P_+ := [P_-^T, s, y]^T \in \mathbb{R}^{(p+2)\times n}$ *and assume that range*$(MP_+) \subset$ *range*(P_+) *Then any solution of* (247) *will have the form*

$$V_+ = \gamma I + P_+^T W_+ P_+$$

for some symmetric matrix $W_+ \in \mathbb{R}^{m+2\times m+2}$.

Remark 5.11. *Before we begin with the proof of Lemma 5.10 let us note, that we will show feasibility of* (247) *(for certain parameter choices) in our situation later on. The reason we have to assume feasibility here is simple: For arbitrary* s, y, μ *and* L *one can easily find choices that make* (247) *infeasible (for example* $y = 2s \neq 0, L = 1$*). Secondly the condition range*$(MP_+) \subset$ *range*(P_+) *is fulfilled by the PSB (Powell symmetric Broyden), DFP (Davidon, Fletcher, Powell) and BFGS (Broyden, Fletcher, Goldfarb, Shanno) weighting matrices* [14]*. Thirdly for small* $p + 2 \ll n$ *we can now focus on an optimization Problem with less variables, namely*

$$\begin{aligned} W_+ \in \underset{W=W^T\in\mathbb{R}^{(p+2)\times(p+2)}}{\textit{argmin}} \quad & ||MP_+^T(W - W_-)P_+M||_F \\ \textit{subject to} \quad & P_+^T W P_+ s = g - \gamma s \\ & \mu I \preceq \gamma I + P_+^T W P_+ \preceq LI \end{aligned} \tag{248}$$

for $W_- := \begin{pmatrix} \hat W_- & 0_{p\times 2} \\ 0_{2\times p} & 0_{2\times 2} \end{pmatrix}$*. Note especially that the semidefinite inequalities can be equivalently stated in the space of symmetric* $(p+2)\times(p+2)$ *matrices. For small* p *we can solve these problems efficiently by for example interior point methods*

Proof. Without loss of generality we can assume that P_+ is of full row rank equal to $p+2$. Now we choose $U \in \mathbb{R}^{(n-p-2)\times n}$ such that $P_+U^T = 0$, $UU^T = I_{n-p-2}$ and $[P_+^T, U^T]$ is invertible. Since $P_+^T, U^T]$ is invertible, we can now write any feasible V of (247) in the form

$$V = \gamma I + [P_+^T, U^T] \begin{pmatrix} W_1 & W_2 \\ W_2^T & W_3 \end{pmatrix} [P_+^T, U^T]^T \tag{249}$$

[14]For BFGS we have to interchange the roles of s and y though.

for some $W_1 = W_1^T \in \mathbb{R}^{(p+2)\times(p+2)}$, $W_2 \in \mathbb{R}^{p\times(n-p-2)}$ and $W_3 \in \mathbb{R}^{(n-p-2)\times(n-p-2)}$. Our first step now is to show that

$$\hat{V} = \gamma I + P_+^T W_1 P_+$$

is also feasible for (247). From $Us = 0$ and $Ug = 0$, we conclude that

$$g = Vs = \gamma s + P_+^T W_1 P_+ s + U^T \underbrace{W_2 P_+ s}_{=0} = \hat{V} s \tag{250}$$

holds true. By multiplying the chain of convex inequalities

$$\mu I \preceq \gamma I + [P_+^T, U^T] \begin{pmatrix} W_1 & W_2 \\ W_2^T & W_3 \end{pmatrix} [P_+^T, U^T]^T \preceq LI \tag{251}$$

with the orthogonal matrix $[P_+^T (P_+ P_+^T)^{-\frac{1}{2}}, U^T]$ from the right and its transpose from the left one obtains the chain of inequalities

$$\mu I \preceq \gamma I + \begin{pmatrix} (P_+P_+^T)^{\frac{1}{2}} W_1 (P_+P_+^T)^{\frac{1}{2}} & (P_+P_+^T)^{\frac{1}{2}} W_2 \\ W_2^T (P_+P_+^T)^{\frac{1}{2}} & W_3 \end{pmatrix} \preceq LI \tag{252}$$

and together with $\mu - \gamma \leq 0$ and $L - \gamma \geq 0$ this implies that

$$\mu I \preceq \gamma I + \begin{pmatrix} (P_+P_+^T)^{\frac{1}{2}} W_1 (P_+P_+^T)^{\frac{1}{2}} & 0 \\ 0 & 0 \end{pmatrix} \preceq LI \tag{253}$$

is fulfilled. If we now finally multiply (253) by $[P_+^T (P_+ P_+^T)^{-\frac{1}{2}}, U^T]$ from the left and with its transpose from the right, we obtain the desired result: $\hat{V}$ is feasible for (247). In our second step will we show that the objective value of (247) does not increase when one replaces V by $\hat{V}$. With the definition of W_- from remark 5.11 we can write $V_- = P_+^T W_- P_+$ and therefore obtain

$$\begin{aligned} ||M(V - V_-)M||_F =& || \begin{pmatrix} P_+ M^2 P_+^T & 0 \\ 0 & I_{n-m-2} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} W_1 - W_- & W_2 \\ W_2^T & W_3 \end{pmatrix} \begin{pmatrix} P_+ M^2 P_+^T & 0 \\ 0 & I_{n-m-2} \end{pmatrix}^{\frac{1}{2}} ||_F \\ \geq& ||(P_+ M^2 P_+^T)^{\frac{1}{2}} (W_1 - W_-)(P_+ M^2 P_+^T)^{\frac{1}{2}} ||_F \\ =& ||M P_+^T (W - W_-) P_+ M||_F \\ =& ||M(\hat{V} - V_-)M|| \end{aligned} \tag{254}$$

where we used the fact that $PMU^T = 0$ as well as some well known trace identities for the first equality. ■

Let us now investigate the applicability for the derivative of the reduced Lagrangian and the generalized absolute value.

5.2.1 NCQNM for the Generalized Absolute Value

Assume again that we are interested in finding a zero of

$$\nabla f(Z) = \tfrac{1}{2}\mathcal{H}[Z] - \tfrac{1}{2}|Z|_{\mathcal{K}} + \mathcal{R} \tag{255}$$

as before. We start by transferring the norm constrained quasi Newton method (NCQNM) approach above to this function. In the Newton iteration (227) we replace the (generalized) second derivative $\mathcal{M}_k$ with an (invertible) approximation $\mathcal{B}_k \approx \mathcal{M}_k \in \partial^2 f(Z)$, i.e the quasi Newton method could take the form

$$Z^{(k+1)} := Z^{(k)} - \mathcal{B}_k^{-1}\nabla f(Z^{(k)}) \tag{256}$$

for $k \in \mathbb{N}_0$ and some initial iterate $Z^{(0)} \in E$. The main questions to be answered is, how can we choose the approximations $\mathcal{B}_k \in \mathcal{L}(E,E)$ such that we can still (hope to) obtain fast local convergence, but at the same time avoid the immense computational cost of the generalized Newton method. Let us translate the usual approach and consider

$$\begin{aligned}\nabla f(Z^{(k+1)}) &= \tfrac{1}{2}\mathcal{H}[Z^{(k)} - \mathcal{B}_k^{-1}\nabla f(Z^{(k)})] - \tfrac{1}{2}|Z^{(k+1)}|_{\mathcal{K}} + \mathcal{R}\\ &= \nabla f(Z^{(k)}) - \tfrac{1}{2}(|Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} + \mathcal{H}[\mathcal{B}_k^{-1}\nabla f(Z^{(k)})])\\ &= \nabla f(Z^{(k)}) - \tfrac{1}{2}(|Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} + (\mathcal{H} - 2\mathcal{B}_k + 2\mathcal{B}_k)[\mathcal{B}_k^{-1}\nabla f(Z^{(k)})])\\ &= -\tfrac{1}{2}(|Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} + (\mathcal{H} - 2\mathcal{B}_k)[\mathcal{B}_k^{-1}\nabla f(Z^{(k)})])\end{aligned} \tag{257}$$

which for $\mathcal{B}_k = \frac{1}{2}(\mathcal{H} - \mathcal{V}_k)$ implies

$$\begin{aligned}\nabla f(Z^{(k)} - \mathcal{B}_k^{-1}\nabla f(Z^{(k)})) &= -\tfrac{1}{2}(|Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} + \mathcal{V}_k[\mathcal{B}_k^{-1}\nabla f(Z^{(k)})]\\ &= -\tfrac{1}{2}(|Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} - \mathcal{V}_k[Z^{(k)} - \mathcal{B}_k^{-1}\nabla f(Z^{(k)}) - Z^{(k)}]\\ &= -\tfrac{1}{2}(|Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} - \mathcal{V}_k[Z^{(k+1)} - Z^{(k)}]\end{aligned} \tag{258}$$

and therefore that in an ideal world we would choose $\mathcal{V}_k$ such that above quantity is equal to zero. Since $\mathcal{B}_k^{-1}$ depends on $\mathcal{V}_k$ this is to hard to achieve in general (as it is equivalent to finding a zero of ∇f). We can however choose the next approximation $\mathcal{V}_{k+1}$ "close" to $\mathcal{V}_k$ from the affine subspace of all linear operators that satisfy the secant equation,

$$\{\mathcal{V} \in \mathcal{L}(E,E) \mid |Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} = \mathcal{V}[Z^{(k+1)} - Z^{(k)}]\} \tag{259}$$

which is non-empty and contains at least one element from $conv(\partial|[Z^{(k)}, Z^{(k+1)}]|_{\mathcal{K}}$ due to the mean value theorem (proposition 2.21). We have already seen that all operators in $\partial|.|_{\mathcal{K}}$ are self-adjoint, and therefore we can restrict the affine subspace above to the affine space of self-adjoint linear operators that satisfy the secant equation,

$$\mathcal{SEC}(Z^{(k)}, Z^{(k+1)}) := \{\mathcal{V} \in \mathcal{L}(E,E) \mid |Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} = \mathcal{V}[Z^{(k+1)} - Z^{(k)}],\ \mathcal{V} = \mathcal{V}^*\} \tag{260}$$

which is therefore again non-empty. In fact note that if $\mathcal{V}_S$ is any self adjoint operator, then the "symmetrization" $\mathcal{V} + \mathcal{V}^*$ of any linear operator $\mathcal{V}$ yields a closer approximation to $\mathcal{V}_S$

in the operator norm

$$\begin{aligned} ||\tfrac{1}{2}(\mathcal{V}+\mathcal{V}^*)-\mathcal{V}_S|| &= ||\tfrac{1}{2}(\mathcal{V}+\mathcal{V}^*)-\tfrac{1}{2}(\mathcal{V}_S+\mathcal{V}_S^*)|| \\ &\leq ||\tfrac{1}{2}(\mathcal{V}-\mathcal{V}_S)||+||\tfrac{1}{2}(\mathcal{V}^*-\mathcal{V}_S^*)|| \\ &= ||\mathcal{V}^*-\mathcal{V}_S^*|| \\ &= ||\mathcal{V}-\mathcal{V}_S|| \end{aligned} \tag{261}$$

than either one of the operators $\mathcal{V}$ or $\mathcal{V}^*$. Now to realize the norm constrained quasi Newton approach we obviously need to specify what exactly we mean by the words "close" and "choose" and also address the computational cost. In order to choose the new linear operator $\mathcal{V}_{k+1}$ we will (again) follow the most common way of projecting the current approximating $\mathcal{V}_k$ onto a (non-empty, closed and convex) subset of $\mathcal{L}(E,E)$ with respect to some (weighted) norm. Note that it is in principle possible to project with respect to the operator norm, but also computationally challenging and therefore unfavorable in practice. In fact we will use a, possibly weighted, Hilbert-Schmidt norm, which is essentially equivalent to the Frobenius-Norm in finite dimension. We denote the Hilbert-Schmidt inner product $\langle \mathcal{X},\mathcal{Y}\rangle_{HS} : \mathcal{L}(E,E)\times\mathcal{L}(E,E)\to\mathbb{R}$ given by

$$\langle \mathcal{X},\mathcal{Y}\rangle_{HS} := Tr(\mathcal{X}^*\mathcal{Y}) \tag{262}$$

where $Tr : \mathcal{L}(E,E)\to\mathbb{R}$ denotes the trace operator $Tr(\mathcal{Z}) = \sum_k \langle \mathcal{Z}[e_k], e_k\rangle$ for some orthonormal Basis $\{e_k\}_k \subset E$ of E. Note that this definition is in fact independent of the orthonormal basis ($Tr(\mathcal{Z})$ is equal to the sum of eigenvalues of $\mathcal{Z}$ which does not rely on any basis representation). Note that $\langle \mathcal{X},\mathcal{Y}\rangle_{HS}$ is clearly bilinear and the equality

$$Tr(\mathcal{Z}) = Tr(\mathcal{Z}^{(*)}) \tag{263}$$

implies symmetry $\langle \mathcal{X},\mathcal{Y}\rangle_{HS} = \langle \mathcal{Y},\mathcal{X}\rangle_{HS}$ as well as positive definiteness of $\langle \mathcal{X},\mathcal{X}\rangle_{HS}$ (i.e. the Hilbert-Schmidt inner product, is really an inner product and $\mathcal{L}(E,E)$ equipped with $\langle .,.\rangle_{HS}$ is a Euclidean space. We denote the induced norm by

$$||\mathcal{X}||_{HS} := \sqrt{\langle \mathcal{X},\mathcal{X}\rangle_{HS}} \tag{264}$$

for $\mathcal{X}\in\mathcal{L}(E,E)$. It is well known, that if we write

$$Sym(E) := \{\mathcal{X}\in\mathcal{L}(E,E) \mid \mathcal{V}=\mathcal{V}^*\} \tag{265}$$

and

$$Skew(E) = \{\mathcal{X}\in\mathcal{L}(E,E) \mid \ \mathcal{V}=-\mathcal{V}^*\} \tag{266}$$

then these spaces are orthogonal $Sym(E)\perp Skew(E)$ (with respect to the Hilbert-Schmidt inner product) and the space

$$\mathcal{L}(E,E) = Sym(E)\oplus Skew(E) \tag{267}$$

decomposes as a direct sum, especially implying that $Sym(E)$ equipped with the Hilbert-Schmidt inner product is again a Euclidean space. We have already seen that the operator norm of every element in the generalized derivative is bounded by one, which especially implies that the eigenvalues of any element of $conv(\partial|[Z^{(k)}, Z^{(k+1)}]|_{\mathcal{K}})$ are contained in the interval $[-1, 1]$. As a consequence the optimization problem

$$\begin{aligned} \underset{\tilde{\mathcal{V}}\in Sym(E)}{\text{minimize}} \quad & ||\mathcal{M}(\tilde{\mathcal{V}} - \mathcal{V}_-)\mathcal{M}||_{HS} \\ \text{subject to} \quad & \tilde{\mathcal{V}}[S] = G \\ & -\mathcal{I} \preceq \tilde{\mathcal{V}} \preceq \mathcal{I} \end{aligned} \tag{268}$$

has a non-empty feasible set for $S = Z^{(k+1)} - Z^{(k)}$ and $G = |Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}}$ independent of k. Therefore, for self-adjoint invertible $\mathcal{M}$ satisfying a range condition as in Lemma 5.10, we know that (268) attains a unique optimal solution. If we now assume that $\mathcal{V}_- = \mathcal{P}^*\tilde{W}_-\mathcal{P}$ satisfies $||V_-|| \leq 1$ and is of low rank , i.e. $p \in \mathbb{N}$ is "small", $\mathcal{P} : E \to \mathbb{R}^p$ is a linear operator, $\mathcal{P}^*$ denotes its adjoint operator and $\tilde{W}_- = \tilde{W}_-^T \in \mathbb{R}^{p\times p}$ is a symmetric matrix, then we know (from Lemma 5.10 and Remark 5.11) that the problem (268) can be reformulated as an SDP of size $(p+2)\times(p+2)$. Specifically, if $\mathcal{P}_+ : E \to \mathbb{R}^{p+2}$ and $W_- = W_-^T \in \mathbb{R}^{(p+2)\times(p+2)}$ denote "extended" operators (for example as in Lemma 5.11.), and $W_+ = W_+^T \in \mathbb{R}^{(p+2)\times(p+2)}$ is an optimal solution of

$$\begin{aligned} \underset{\tilde{W}=\tilde{W}^T\in\mathbb{R}^{(p+2)\times(p+2)}}{\text{minimize}} \quad & ||(\mathcal{P}_+\mathcal{M}^2\mathcal{P}_+^*)^{\frac{1}{2}}(\tilde{W} - W_-)(\mathcal{P}_+\mathcal{M}^2\mathcal{P}_+^*)^{\frac{1}{2}}||_{HS} \\ \text{subject to} \quad & \mathcal{P}_+^*\tilde{W}\mathcal{P}_+[S] = G \\ & -\mathcal{I} \preceq (\mathcal{P}_+\mathcal{P}_+^*)^{\frac{1}{2}}\tilde{W}(\mathcal{P}_+\mathcal{P}_+^*)^{\frac{1}{2}} \preceq \mathcal{I} \end{aligned} \tag{269}$$

then $V_+ := \mathcal{P}_+^*\mathcal{W}_+\mathcal{P}_+$ solves (268). Whenever p is relatively small, we can solve (269) efficiently by iterative methods. Specifically we will use the PSB weight matrix

$$\mathcal{M} := I \tag{270}$$

in our numerical implementation, as this choice is always well defined and easy to implement. We will address further implementation details later on, especially how to recursively form the product $\mathcal{P}_+\mathcal{P}_+^*$. For now let us assume that this product is cheaply available and let us point out that that if $\mathcal{M}^2$ is equal to the sum of the identity and a low rank matrix, then $\mathcal{P}\mathcal{M}^2\mathcal{P}^*$ might also be cheaply available. This is for example the case for the DFP and BFGS weighting matrices, although they need to be modified[15] whenever the inner product $\langle S, G\rangle < 0$ is negative and can not be defined when the inner product $\langle S, G\rangle = 0$ is equal to zero (as this would contradict positive definiteness of $\mathcal{M}^2$, assuming that $G \neq 0 \neq S$).

Note that our work so far, can be easily transfered to the broader setting of locally Lipschitz continuous gradients, whenever we know something about the (local) constants. In the next subsection we will investigate a certain structure of the generalized absolute value (which is not necessarily shared by the class of Lipschitz continuous gradients), that

[15] For example we can choose invertible $M = M^T$ conditionally such that

$$\begin{aligned} g^T s > 0 \Rightarrow M^2 g = s \\ g^T s < 0 \Rightarrow M^2 g = -s \end{aligned} \tag{271}$$

is satisfied.

allows a certain local refinement. It turns out that our refinement exploiting this structure will be in conflict with the full exploitation of certain averaged (or global) information, but nevertheless seems to work well in our numerical examples.

5.2.2 Local Refinement

Our local refinement is based on the following observation. Although we do not know much about the eigenvectors (apart from their existence) of $\frac{1}{2}(\mathcal{H} - \mathcal{V}_B) \in \partial_B^2 f(Z)$ we have access to up to two eigenvectors of $\mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}}$, and their computation involves almost no extra computational cost.

Proposition 5.12. *Let $Z \in E$ be arbitrary. Then the equalities*

$$\mathcal{V}_B[\Pi_{\mathcal{K}}(Z)] = \Pi_{\mathcal{K}}(Z) \; \forall \; \mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}} \tag{272}$$

and

$$\mathcal{V}_B[\Pi_{\mathcal{K}^P}(Z)] = -\Pi_{\mathcal{K}^P}(Z) \; \forall \; \mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}} \tag{273}$$

hold true, i.e. whenever the projections are not zero ($\Pi_{\mathcal{K}}(Z) \neq 0$ or $\Pi_{\mathcal{K}^P}(Z) \neq 0$) they are eigenvectors (with eigenvalues one respectively minus one) of all elements $\mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}}$ and therefore of all elements $\mathcal{V} \in \partial |Z|_{\mathcal{K}}$.

Remark 5.13. *For $Z \in E$ let us define the affine subspace*

$$\mathcal{EIG}(Z) := \{\mathcal{V} \in \mathcal{L}(E, E) \mid \mathcal{V} = \mathcal{V}^*, \mathcal{V}[\Pi_{\mathcal{K}}(Z)] = \Pi_{\mathcal{K}}(Z), \; \mathcal{V}[\Pi_{\mathcal{K}^P}(Z)] = -\Pi_{\mathcal{K}^P}(Z)\} \tag{274}$$

and note that (7) *readily provides a formula to project (with respect to any induced norm) onto $\mathcal{EIG}(Z)$.*

Proof of Proposition 5.12. Let $\mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}}$ be arbitrary. Each equality is trivial whenever the projections are equal to zero. Let us therefore assume that $\Pi_{\mathcal{K}}(Z) \neq 0$ and $\Pi_{\mathcal{K}^P}(Z) \neq 0$. By definition of the B-derivative there exists a sequence $\{Z_k\}_{k \in \mathbb{N}} \subset E \setminus \Omega_{\nabla f}$ with $\lim_{k \to \infty} Z_k = Z$ and $\lim_{k \to \infty} \nabla |Z_k|_{\mathcal{K}} = \mathcal{V}_B$. Let us assume without loss of generality that $\Pi_{\mathcal{K}}(Z_k)$ and $\Pi_{\mathcal{K}^P}(Z_k)$ are not equal to zero for all $k \in \mathbb{N}$. If we replace Z with Z_k and $\Delta Z = t\Pi_{\mathcal{K}}(Z_k)$ in the definition of F-differentiability we obtain

$$\begin{aligned}
0 &= \lim_{t \to 0} \frac{\| |Z_k + t\Pi_{\mathcal{K}}(Z_k)|_{\mathcal{K}} - |Z_k|_{\mathcal{K}} - \nabla |Z_k|_{\mathcal{K}}[t\Pi_{\mathcal{K}}(Z_k)] \|}{\| t\Pi_{\mathcal{K}}(Z_k) \|} \\
&= \lim_{t \to 0} \frac{\| |(1+t)\Pi_{\mathcal{K}}(Z_k) + \Pi_{\mathcal{K}^P}(Z_k)|_{\mathcal{K}} - |Z_k|_{\mathcal{K}} - \nabla |Z_k|_{\mathcal{K}}[t\Pi_{\mathcal{K}}(Z_k)] \|}{\| t\Pi_{\mathcal{K}}(Z_k) \|} \\
&= \lim_{t \to 0} \frac{\| (1+t)\Pi_{\mathcal{K}}(Z_k) - \Pi_{\mathcal{K}^P}(Z_k) - |Z_k|_{\mathcal{K}} - \nabla |Z_k|_{\mathcal{K}}[t\Pi_{\mathcal{K}}(Z_k)] \|}{\| t\Pi_{\mathcal{K}}(Z_k) \|} \\
&= \lim_{t \to 0} \frac{\| t\Pi_{\mathcal{K}}(Z_k) - \nabla |Z_k|_{\mathcal{K}}[t\Pi_{\mathcal{K}}(Z_k)] \|}{\| t\Pi_{\mathcal{K}}(Z_k) \|} \\
&= \lim_{t \to 0} \frac{\| \Pi_{\mathcal{K}}(Z_k) - \nabla |Z_k|_{\mathcal{K}}[\Pi_{\mathcal{K}}(Z_k)] \|}{\| \Pi_{\mathcal{K}}(Z_k) \|} \\
&= \frac{\| \Pi_{\mathcal{K}}(Z_k) - \nabla |Z_k|_{\mathcal{K}}[\Pi_{\mathcal{K}}(Z_k)] \|}{\| \Pi_{\mathcal{K}}(Z_k) \|}
\end{aligned} \tag{275}$$

which implies $\nabla|Z_k|_{\mathcal{K}}[\Pi_{\mathcal{K}}(Z_k)] = \Pi_{\mathcal{K}}(Z_k)\ \forall k \in \mathbb{N}$. This proves the equality of our claim, since all terms in this equation are convergent for $k \to \infty$. The second equality follows by replacing Z with Z_k and $\Delta Z = -t\Pi_{\mathcal{K}^P}(Z_k)$ in the definition of F-differentiability. We obtain

$$\begin{aligned}
0 &= \lim_{t\to 0} \frac{\| |Z_k - t\Pi_{\mathcal{K}^P}(Z_k)|_{\mathcal{K}} - |Z_k|_{\mathcal{K}} - \nabla|Z_k|_{\mathcal{K}}[-t\Pi_{\mathcal{K}^P}(Z_k)]\|}{\|t\Pi_{\mathcal{K}^P}(Z_k)\|} \\
&= \lim_{t\to 0} \frac{\| |\Pi_{\mathcal{K}^P}(Z_k) + (1-t)\Pi_{\mathcal{K}^P}(Z_k)|_{\mathcal{K}} - |Z_k|_{\mathcal{K}} + \nabla|Z_k|_{\mathcal{K}}[t\Pi_{\mathcal{K}^P}(Z_k)]\|}{\|t\Pi_{\mathcal{K}^P}(Z_k)\|} \\
&= \lim_{t\to 0} \frac{\|\Pi_{\mathcal{K}}(Z_k) - (1-t)\Pi_{\mathcal{K}^P}(Z_k) - |Z_k|_{\mathcal{K}} + \nabla|Z_k|_{\mathcal{K}}[t\Pi_{\mathcal{K}^P}(Z_k)]\|}{\|t\Pi_{\mathcal{K}^P}(Z_k)\|} \\
&= \lim_{t\to 0} \frac{\|t\Pi_{\mathcal{K}^P}(Z_k) + \nabla|Z_k|_{\mathcal{K}}[t\Pi_{\mathcal{K}^P}(Z_k)]\|}{\|t\Pi_{\mathcal{K}^P}(Z_k)\|} \\
&= \lim_{t\to 0} \frac{\|\Pi_{\mathcal{K}^P}(Z_k) + \nabla|Z_k|_{\mathcal{K}}[\Pi_{\mathcal{K}^P}(Z_k)]\|}{\|\Pi_{\mathcal{K}^P}(Z_k)\|} \\
&= \frac{\|\Pi_{\mathcal{K}^P}(Z_k) + \nabla|Z_k|_{\mathcal{K}}[\Pi_{\mathcal{K}^P}(Z_k)]\|}{\|\Pi_{\mathcal{K}^P}(Z_k)\|}
\end{aligned} \tag{276}$$

and therefore $\nabla|Z_k|_{\mathcal{K}}[\Pi_{\mathcal{K}^P}(Z_k)] = -\Pi_{\mathcal{K}^P}(Z_k)$. Again our claim follows by noting that all terms are convergent for $k \to \infty$. ■

Remark 5.14. *Let us also give an example of a possible further refinement for a certain block structure. Assume that our cone $\mathcal{K} \subset E$ and our space E are given as Cartesian products, i.e. $\mathcal{K}$ is given as a Cartesian product of two (non-empty, closed and convex) cones $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2 \subset E_1 \times E_2 = E$ for some Euclidean spaces E_1, E_2 satisfying $\langle .,. \rangle_E = \langle .,. \rangle_{E_1} + \langle .,. \rangle_{E_2}$. Then we obtain for $Z = (Z_1, Z_2) \in E$ the decomposition $\Pi_{\mathcal{K}}(Z) = (\Pi_{\mathcal{K}_1}(Z_1), \Pi_{\mathcal{K}_2}(Z_2)) = \underbrace{(\Pi_{\mathcal{K}_1}(Z_1), 0)}_{\in\mathcal{K}} + \underbrace{(0, \Pi_{\mathcal{K}_2}(Z_2))}_{\in\mathcal{K}}$ and by transferring the limit argument from our previous proof, we see that $X_1 = (\Pi_{\mathcal{K}_1}(Z_1), 0), X_2 = (0, \Pi_{\mathcal{K}_2}(Z_2))$ must also satisfy the equation $\mathcal{V}_B[X_i] = X_i$, for $i \in \{1, 2\}$.*

By linear combination, we see that knowledge about the generalized derivative implies knowledge about the value of the generalized absolute value. This means that evaluating the generalized derivative is always at least as "difficult" as evaluating the function itself:

Corollary 5.15. *Let $Z \in E$ be arbitrary. Then the equalities*

$$\mathcal{V}_B[Z] = |Z|_{\mathcal{K}}\ \forall\ \mathcal{V}_B \in \partial_B|Z|_{\mathcal{K}} \tag{277}$$

and

$$\mathcal{V}_B[|Z|_{\mathcal{K}}] = Z\ \forall\ \mathcal{V}_B \in \partial_B|Z|_{\mathcal{K}} \tag{278}$$

hold true.

Proof. By applying proposition 5.10 we see that $\mathcal{V}_B[Z] = \mathcal{V}_B[\Pi_{\mathcal{K}}(Z) + \Pi_{\mathcal{K}^P}(Z)] = \Pi_{\mathcal{K}}(Z) - \Pi_{\mathcal{K}^P}(Z) = |Z|_{\mathcal{K}}$ and $\mathcal{V}_B[|Z|_{\mathcal{K}}] = \mathcal{V}_B[\Pi_{\mathcal{K}}(Z) - \Pi_{\mathcal{K}^P}(Z)] = \Pi_{\mathcal{K}}(Z) + \Pi_{\mathcal{K}^P}(Z) = Z$ hold true for any $\mathcal{V}_B \in \partial_B|Z|_{\mathcal{K}}$. ■

Corollary 5.16. *Let $Z \in E \setminus \{0\}$ be arbitrary. Then the operator norm of the generalized derivatives at Z is always equal to one, i.e.*

$$||\mathcal{V}_B|| = 1 \ \forall \ \mathcal{V}_B \in \partial_B |Z|_{\mathcal{K}} \tag{279}$$

and

$$||\mathcal{V}|| = 1 \ \forall \ \mathcal{V} \in \partial |Z|_{\mathcal{K}} \tag{280}$$

hold true.

Proof. We have already seen that the inequalities $||\mathcal{V}_B|| \leq 1$ and $||\mathcal{V}|| \leq 1$ are true. Since $Z \neq 0$ we obtain that $\Pi_{\mathcal{K}}(Z)$ and $\Pi_{\mathcal{K}^P}(Z)$ can not be both equal to zero. With the convention $\frac{0}{0} = 0$ we can use the operator norm definition in conjunction with proposition 5.10 to show

$$\begin{aligned} 1 \geq ||\mathcal{V}|| &= \sup_{X \in E} \{\|\mathcal{V}[Z]\| \ | \ \|X\| = 1\} \\ &\geq \max\{\left\|\mathcal{V}[\tfrac{\Pi_{\mathcal{K}}(Z)}{\|\Pi_{\mathcal{K}}(Z)\|}]\right\|, \left\|\mathcal{V}[\tfrac{\Pi_{\mathcal{K}^P}(Z)}{\|\Pi_{\mathcal{K}^P}(Z)\|}]\right\|\} \\ &= \max\{\left\|\tfrac{\Pi_{\mathcal{K}}(Z)}{\|\Pi_{\mathcal{K}}(Z)\|}\right\|, \left\|\tfrac{\Pi_{\mathcal{K}^P}(Z)}{\|\Pi_{\mathcal{K}^P}(Z)\|}\right\|\} = 1 \end{aligned} \tag{281}$$

and therefore $||\mathcal{V}|| = 1 \ \forall \mathcal{V} \in \partial |Z|_{\mathcal{K}}$. ■

Let us define the unit sphere

$$S_{\mathcal{L}} := \{\mathcal{V} \in \mathcal{L}(E, E) \mid \mathcal{V} = \mathcal{V}^*, \ ||\mathcal{V}|| = 1\} \tag{282}$$

of self-adjoint linear operators with norm equal to one and the unit ball

$$B_{\mathcal{L}} := \{\mathcal{V} \in \mathcal{L}(E, E) \mid \mathcal{V} = \mathcal{V}^*, \ ||\mathcal{V}|| \leq 1\} \tag{283}$$

of self-adjoint linear operators with norm smaller or equal to one. Then another way of stating corollary 5.15 is that

$$\partial |Z|_{\mathcal{K}} \subset S_{\mathcal{L}} \ \forall Z \in E \setminus \{0\} \tag{284}$$

the generalized derivatives are contained in the unit sphere. Obviously we would like to exploit the obtained equalities somehow for our quasi Newton method, but as one would expect there is a compatibility issue. Recall the mean value theorem

$$|Z^{(k+1)}|_{\mathcal{K}} - |Z^{(k)}|_{\mathcal{K}} \in conv(\partial |[Z^{(k)}, Z^{(k+1)}]|_{\mathcal{K}})[Z^{(k+1)} - Z^{(k)}] \tag{285}$$

for the generalized absolute value. Even though the "endpoints" $\partial |Z^{(k)}|_{\mathcal{K}}$ and $\partial |Z^{(k+1)}|_{\mathcal{K}}$ are both contained in the unit sphere $S_{\mathcal{L}}$ whenever $Z^{(k)} \neq 0 \neq Z^{(k+1)}$, there is no implication for the set $\partial |[Z^{(k)}, Z^{(k+1)}]|_{\mathcal{K}}$ other than it being fully contained inside the unit ball $B_{\mathcal{L}}$.

Example 5.17. *Consider the absolute value $|.|_{\mathbb{R}_+} : \mathbb{R} \to \mathbb{R}$ and let $Z := -1, \Delta Z = 2$. Then $|Z|_{\mathbb{R}_+} = 1, |Z + \Delta Z|_{\mathbb{R}_+} = 1, \partial |Z|_{\mathbb{R}_+} = -1$, and $\partial |Z + \Delta Z|_{\mathbb{R}_+} = 1$ holds true. On the other hand $0 = 1 - 1 = |Z + \Delta Z|_{\mathbb{R}_+} - |Z|_{\mathbb{R}_+} \in \partial |[Z, Z + \Delta Z]|_{\mathbb{R}_+}[Z + \Delta Z - Z] = \partial |[-1, 1]|_{\mathbb{R}_+}[1]$ implies that the unique element from $\partial |[Z, Z + \Delta Z]|_{\mathbb{R}_+}$ that satisfies the secant equation is given by $\mathcal{V} = 0$ and its operator norm is clearly smaller than one.*

On the other hand, consider a point $Z \in E$ and some self-adjoint operator $\mathcal{V} = \mathcal{V}^* \in \mathcal{L}(E, E)$, we consider now the optimization problem

$$\begin{aligned} \Pi_{\mathcal{EIG}(Z)}(\mathcal{V}) := \underset{\tilde{\mathcal{V}} \in \mathcal{L}(E,E), \tilde{\mathcal{V}}^* = \tilde{\mathcal{V}}}{\text{argmin}} \quad & ||\tilde{\mathcal{V}} - \mathcal{V}||_{HS} \\ \text{subject to} \quad & \tilde{\mathcal{V}}[\Pi_{\mathcal{K}}(Z)] = \Pi_{\mathcal{K}}(Z) \\ & \tilde{\mathcal{V}}[\Pi_{\mathcal{K}^P}(Z)] = -\Pi_{\mathcal{K}^P}(Z) \end{aligned} \tag{286}$$

which we can solve by using (7) (where we replace E by $Sym(E)$ and $\langle ., . \rangle$ by $\langle ., . \rangle_{HS}$. Let us give an explicit formula for $E = \mathbb{R}^n$ and the standard scalar-product (which is nicer to write down and one can easily adapt the formula for the general case). Note that in this case the Hilbert-Schmidt and Frobenius norm coincide. To keep our notation short we write x and y in place of $\Pi_{\mathcal{K}}(Z)$ and $\Pi_{\mathcal{K}^P}(Z)$.

Lemma 5.18. *(Eigen-Update) Let $V = V^T \in \mathbb{R}^{n \times n}$, $x, y \in \mathbb{R}^n$ with $x^T y = 0$. Then the unique solution of the problem*

$$\begin{aligned} \underset{\tilde{V} = \tilde{V}^T \in \mathbb{R}^{n \times n}}{\textit{minimize}} \quad & ||\tilde{V} - V||_F \\ \textit{subject to} \quad & \tilde{V}x = x \\ & \tilde{V}y = -y \end{aligned} \tag{287}$$

is given by

$$V_{Eig} := (I_n - \tfrac{xx^T}{||x||_2^2} - \tfrac{yy^T}{||y||_2^2})V(I_n - \tfrac{xx^T}{||x||_2^2} - \tfrac{yy^T}{||y||_2^2}) + \tfrac{xx^T}{||x||_2^2} - \tfrac{yy^T}{||y||_2^2} \tag{288}$$

where $I_n \in \mathbb{R}^{n \times n}$ denotes the identity. We use the convention $\frac{0}{0} = 0$, in case x, y or both are zero.

Proof. This can be easily derived from the projection formula for projecting onto affine subspaces. ■

Corollary 5.19. *Let $V = V^T \in \mathbb{R}^{n \times n}$ with $||V|| \leq 1$ be satisfied and x, y, V_{Eig} as in the previous lemma, then we obtain*

$$||V_{Eig}|| = 1 \tag{289}$$

if at least one of the vectors x and y is not equal to zero (If they are both equal to zero we obtain $V_{Eig} = V$).

Proof. If at least one of the vectors x, y is non-zero then V_{Eig} has one eigenvalue with absolute value equal to one. We immediately conclude that $||V_{Eig}|| \geq 1$ must be satisfied. Let us for simplicity assume $x \neq 0 \neq y$. Since V_{Eig} is symmetric we can find $n-2$ linear independent Eigenvectors $\{u_i\}_{i=1..n-2}$ of V_{Eig} (with corresponding eigenvalues λ_i) that are orthogonal to x and y. We obtain

$$|\lambda_i| = |\tfrac{u_i^T V_{Eig} u_i}{||u_i||^2}| = |\tfrac{u_i^T V u_i}{||u_i||^2}| \leq ||V|| \leq 1 \tag{290}$$

and therefore $||V_{Eig}|| = \max\{1, |\lambda_1|, ..., |\lambda_{n-2}|\} \leq 1$, which proves our claim. ■

Coming back to our general notation, we see that for $\mathcal{V} \in Sym(E)$ with $||\mathcal{V}|| \leq 1$ the equality

$$\Pi_{\mathcal{EIG}(Z)\cap B_{\mathcal{L}}}(\mathcal{V}) = \Pi_{\mathcal{EIG}(Z)}(\mathcal{V}) \tag{291}$$

and for $Z \neq 0$

$$\Pi_{\mathcal{EIG}(Z)\cap S_{\mathcal{L}}}(\mathcal{V}) = \Pi_{\mathcal{EIG}(Z)}(\mathcal{V}) \tag{292}$$

holds true. Let us summarize what we have learned so far. Assume that we are given $Z^{(k)}, Z^{(k+1)} \in E$ and $\mathcal{V}_k \in \mathcal{L}(E,E)$ with $\mathcal{V}_k = \mathcal{V}_k^*$.

- The mean value theorem might yield relevant averaged curvature information, which we can exploit by trying to satisfy the secant equation while taking into account that the operator should be self-adjoint (260) and its norm smaller than or equal to one, cf. (283). In formulas
$$\mathcal{V}_{k+\frac{1}{2}} \in \mathcal{SEC}(Z^{(k)}, Z^{(k+1)}) \cap B_{\mathcal{L}}. \tag{293}$$

- Provided that $Z^{(k+1)} \neq 0$, the new approximation should be self-adjoint, have operator norm equal to one and the projections $\Pi_{\mathcal{K}}(Z^{(k+1)})$ and $\Pi_{\mathcal{K}^P}(Z^{(k+1)})$ should be eigenvectors with eigenvalues one or minus one (whenever these projections are not equal to zero). In formulas
$$\mathcal{V}_{k+1} \in \mathcal{EIG}(Z^{(k+1)}) \cap S_{\mathcal{L}} \tag{294}$$
is what we want. Note that the equality $\mathcal{EIG}(Z^{(k+1)}) \cap S_{\mathcal{L}} = \mathcal{EIG}(Z^{(k+1)}) \cap B_{\mathcal{L}}$ is true whenever $Z^{(k+1)} \neq 0$.

- In the following we need to investigate realizations that are also computationally affordable. We will essentially define the new approximation as projections
$$\mathcal{V}_{++} := \Pi_{\mathcal{EIG}(Z^{(k+1)})\cap B_{\mathcal{L}}}(\Pi^{\mathcal{M}_k}_{\mathcal{SEC}(Z^{(k)},Z^{(k+1)})\cap B_{\mathcal{L}}}(\mathcal{V}_k)) \tag{295}$$
of the old ones with respect to some induced (weighted) norms. Specifically we can control (up to a point) the rank of the new approximation, which motivates a limited memory strategy, i.e. for an a priori fixed natural number $r \leq dim(E)$ we will use some "dementia" functions $\mathcal{D}_k : \mathcal{L}(E,E) \to \mathcal{L}(E,E)$ that limits $rank(\mathcal{D}_k(\mathcal{V}_k)) \leq 2r$ but also respects the operator norm $||\mathcal{D}_k(\mathcal{V}_k)|| \leq 1$ and self-adjointness $\mathcal{D}_k(V_k))^* = (\mathcal{D}_k(V_k)$. Our final update procedure is then given by
$$\mathcal{V}_{k+1} := \Pi_{\mathcal{EIG}(Z^{(k+1)})\cap B_{\mathcal{L}}}(\Pi^{\mathcal{M}_k}_{\mathcal{SEC}(Z^{(k)},Z^{(k+1)})\cap B_{\mathcal{L}}}(\mathcal{D}_k(\mathcal{V}_k))) \tag{296}$$
and can be used in combination with the Shermann-Morrison formula to make the inversion of $\frac{1}{2}(\mathcal{H} - \mathcal{V}_{k+1})$ relatively cheap for small r. The overall additional memory requirement is upper bounded by $2r(\dim(E) + m + 6r)$. The number of additional flops per iteration will be in $O(\dim(E)r + r^3)$ plus essentially one additional evaluation of $\mathcal{H}[Z]$.

5.2.3 Recursive Inversion

Let us continue by adding one further ingredient in order to make our norm constrained quasi Newton method efficient. We need to be able to "cheaply" compute the directions $\Delta Z^{(k)} = 2(\mathcal{H} - \mathcal{V}_k)^{-1}\nabla f(Z^{(k)})$. We will do so, by ensuring that our approximation $\mathcal{V}_k = \mathcal{P}_k^*\mathcal{W}_k\mathcal{P}_k$ are self-adjoint, low rank and then employ the following corollary from the well known Woodbury matrix identity in combination with a recursion (in k).

Corollary 5.20. *Let $\mathcal{H} : E \to E$ be an orthogonal and self-adjoint linear operator. Let $\mathcal{V} : E \to E$ a linear self-adjoint operator with a decomposition $\mathcal{V} = \mathcal{P}^*\mathcal{W}\mathcal{P}$ such that $\mathcal{W}$ is invertible. If $\mathcal{W}^{-1} - \mathcal{P}\mathcal{H}\mathcal{P}^*$ is non-singular, then*

$$(\mathcal{H} - \mathcal{V})^{-1} = \mathcal{H} + \mathcal{H}\mathcal{P}^*(\mathcal{W}^{-1} - \mathcal{P}\mathcal{H}\mathcal{P}^*)^{-1}\mathcal{P}\mathcal{H} \tag{297}$$

Proof. This is a direct implication from the well known Woodsbury matrix identity (see [94]). ∎

Let us for simplicity assume that $\mathcal{V}_+$ is given as $\mathcal{V}_+ = \mathcal{P}_+^* W_+ \mathcal{P}_+$ for linear $\mathcal{P}_+ : E \to \mathbb{R}^{2p+2}$ and invertible $W_+ \in \mathbb{R}^{(2p+2)\times(2p+2)}$ such that $W^{-1} - \mathcal{P}_+\mathcal{H}\mathcal{P}_+^* \in \mathbb{R}^{2(p+2)\times 2(p+2)}$ is invertible. In order to really exploit the Woodsbury formula we employ a "trick". Assuming that the linear operator $\mathcal{P}_+ : E \to \mathbb{R}^{2p+2}$ is a combination of an "old" (known) linear operator $\mathcal{P} : E \to \mathbb{R}^{2p}$ and new information, say $X, Y \in E$, it is reasonable to assume that $\Phi := \mathcal{P}\mathcal{P}^*$ and $\Psi := \mathcal{P}\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}\mathcal{P}^*$ have been computed beforehand. Then the new quantities $\Phi_+ := \mathcal{P}_+\mathcal{P}_+^*$ and $\Psi_+ := \mathcal{P}_+\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}\mathcal{P}_+^*$ can be computed relatively cheaply from the old quantities Φ and Ψ. For example for

$$\mathcal{P}_+ = \begin{pmatrix} \mathcal{P}[.] \\ \langle X, . \rangle \\ \langle Y, . \rangle \end{pmatrix} \tag{298}$$

we obtain

$$\Phi_+ = \mathcal{P}_+\mathcal{P}_+^* = \begin{pmatrix} \mathcal{P}\mathcal{P}^* & \mathcal{P}[X] & \mathcal{P}[Y] \\ \mathcal{P}[X]^T & \|X\|^2 & \langle X, Y\rangle \\ \mathcal{P}[Y]^T & \langle Y, X\rangle & \|Y\|^2 \end{pmatrix} = \begin{pmatrix} \Phi & \mathcal{P}[X] & \mathcal{P}[Y] \\ \mathcal{P}[X]^T & \|X\|^2 & \langle X, Y\rangle \\ \mathcal{P}[Y]^T & \langle Y, X\rangle & \|Y\|^2 \end{pmatrix} \tag{299}$$

and similar

$$\begin{aligned} \Psi_+ =& \mathcal{P}_+\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}\mathcal{P}_+^* \\ =& \begin{pmatrix} \Psi & \mathcal{P}[\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(X)] & \mathcal{P}[\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(Y)] \\ \mathcal{P}[\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(X)]^T & \left\|\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(X)\right\|^2 & \langle \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(X), Y\rangle \\ \mathcal{P}[\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(Y)]^T & \langle \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(Y), X\rangle & \left\|\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(Y)\right\|^2 \end{pmatrix} \end{aligned} \tag{300}$$

which can be explicitly computed with two evaluations of $\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}$ four evaluations of $\mathcal{P}$ and no more then six additional evaluations of $\langle .,. \rangle$. The formula

$$\begin{aligned} \tfrac{1}{2}\Delta Z :=& (\mathcal{H} - \mathcal{V}_+)^{-1}\nabla f(Z) \\ =& (\mathcal{H} + \mathcal{H}\mathcal{P}_+^*(\mathcal{W}_+^{-1} - \mathcal{P}_+\mathcal{H}\mathcal{P}_+^*)^{-1}\mathcal{P}_+\mathcal{H})\nabla f(Z) \\ =& (\mathcal{I} + \mathcal{H}\mathcal{P}_+^*(\mathcal{W}_+^{-1} - \mathcal{P}_+\mathcal{H}\mathcal{P}_+^*)^{-1}\mathcal{P}_+)\mathcal{H}\nabla f(Z) \\ =& (\mathcal{I} + \mathcal{H}\mathcal{P}_+^*(\mathcal{W}_+^{-1} - \mathcal{P}_+(\mathcal{I} - 2\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})\mathcal{P}_+^*)^{-1}\mathcal{P}_+)\mathcal{H}\nabla f(Z) \\ =& (\mathcal{I} + \mathcal{H}\mathcal{P}_+^*\underbrace{(\mathcal{W}_+^{-1} - \Phi_+ + 2\Psi_+)}_{\in\mathbb{R}^{2(p+1)\times 2(p+1)}}{}^{-1}\mathcal{P}_+)\mathcal{H}\nabla f(Z) \end{aligned} \tag{301}$$

then implies that we can derive the directions ΔZ from $\mathcal{H}[\nabla f(Z)]$ with one application of $\mathcal{P}^*$, one of $\mathcal{P}$ and one of $\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}$. Now depending on the evaluation cost of $\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}$ it might make sense to trade reduced computational effort for additional storage requirements. If we are willing to save $\Theta := (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}\mathcal{P}^*$ then

$$\begin{aligned} \Theta_+ :=& (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}\mathcal{P}_+^* \\ =& \Big((\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}\mathcal{P}^*, (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(X), (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(Y)\Big) \\ =& \Big(\Theta, (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(X), (\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(Y)\Big) \end{aligned} \tag{302}$$

can be used to derive

$$\begin{aligned} \tfrac{1}{2}\Delta Z :=& (\mathcal{H} - \mathcal{V}_+)^{-1}\nabla f(Z) \\ =& (\mathcal{I} + (\mathcal{P}_+^* - 2\mathcal{A}^*\Theta_+)\underbrace{(\mathcal{W}_+^{-1} - \Phi_+ + 2\Psi_+)}_{\in\mathbb{R}^{2(p+1)\times 2(p+1)}}{}^{-1}\mathcal{P}_+)\mathcal{H}\nabla f(Z) \end{aligned} \tag{303}$$

which might reduce overall computational time. In fact a procedure of the following type

1. Input: $X := \Pi_{\mathcal{K}}(Z), Y := \Pi_{\mathcal{K}^P}(Z), \mathcal{P}, \Phi, \Psi, \Theta, W$
2. Output: $\mathcal{H}\nabla f(Z)$ and $\Delta Z := \frac{1}{2}(\mathcal{H} - \mathcal{P}_+^*\mathcal{W}_+\mathcal{P}_+)^{-1}\nabla f(Z)$ as well as $\Phi_+, \Psi_+, \Theta_+, W_+$

can be realized where $(\mathcal{A}\mathcal{A}^*)^{-1}$ has to be evalutated only twice. Note that the evaluation of $\mathcal{H}\nabla f(Z)$ readily requires one evaluation of $(\mathcal{A}\mathcal{A}^*)^{-1}$.

5.2.4 Limited Memory via Compression

The recursive inversion strategy of the last section is only meaningful if we limit the rank of our approximations $\mathcal{V}_k \in Sym(E)$. Let us assume we want to enforce the condition $rank(\mathcal{V}_k) \leq 2r \ \forall k \in \mathbb{N}$ and some $r \in \mathbb{N}$ and that we are given a representation

$$\mathcal{V}_- = \mathcal{P}_-^*\mathcal{W}_-\mathcal{P}_- \tag{304}$$

for $\mathcal{P}_- : E \to \mathbb{R}^{2p}$, symmetric $W_- \in \mathbb{R}^{(2p)\times(2p)}$. If $p < r$ holds true then we can easily append new information as in (298) (with $\mathcal{P}_-$ instead of $\mathcal{P}$, and any $X, Y \in E$) without risking to violate the condition $rank(\mathcal{V}_+) \leq 2r$. In contrast if $p \geq r$ is satisfied, then any appending, e.g. as in (298), is likely to violate the rank condition. The simplest way, that comes to mind, to deal with this inconvenience is to simply delete old information. For example if the linear operator is given as

$$\mathcal{P}_- = \begin{pmatrix} \langle X_{old}, .\rangle \\ \langle Y_{old}, .\rangle \\ \mathcal{P} \end{pmatrix} \tag{305}$$

then we could simply work with $\mathcal{P}$ instead of $\mathcal{P}_-$ as well as the corresponding submatrix $W \in \mathbb{R}^{2(p-1)\times 2(p-1)}$ of W_-. Unfortunately the deletion of the first two "rows" of $\mathcal{P}_-$ might yield

$$||\mathcal{P}^* W \mathcal{P}|| > 1 \tag{306}$$

even if the operator norm of V_- is bounded from above by one.

Example 5.21. *Consider the matrix* $V_- = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$ *which has operator (spectral) norm equal to one (its eigenvalues are plus and minus one). We can write* $V_- = P_-^T W_- P_-$ *for* $P_- = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ *and* $W_- = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & -\sqrt{2} \end{pmatrix}$*. If we delete the first row of* P_- *and work only with the corresponding submatrix of* W_-*, we obtain the product* $\begin{pmatrix} 0 & 1 \end{pmatrix}^T (W_-)_{22} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -\sqrt{2} \end{pmatrix}$ *which has spectral norm* $\sqrt{2} > 1$*.*

The obvious reason for this unwanted behavior can of course be traced back to the (possible) non-orthogonality of the "rows" of $\mathcal{P}_-$. For simplicity let us assume that $\mathcal{P}_-$ is surjective for now. We could then orthogonalize the "rows" of $\mathcal{P}_-$ by using the inverse symmetric square root $(\mathcal{P}_-\mathcal{P}_-^*)^{-\frac{1}{2}}$. In fact we could rewrite

$$V_- = \mathcal{P}_-^* W_- \mathcal{P}_- = \mathcal{P}_-^*(\mathcal{P}_-\mathcal{P}_-^*)^{-\frac{1}{2}}(\mathcal{P}_-\mathcal{P}_-^*)^{-\frac{1}{2}} W_- (\mathcal{P}_-\mathcal{P}_-^*)^{\frac{1}{2}}(\mathcal{P}_-\mathcal{P}_-^*)^{-\frac{1}{2}}\mathcal{P}^* \tag{307}$$

and use our deleting strategy on $(\mathcal{P}_-\mathcal{P}_-^*)^{-\frac{1}{2}}$ (instead of $\mathcal{P}_-$) without risking to violate our operator norm condition (whenever $||V_-|| \leq 1$ is satisfied). We will avoid this approach because it does not necessarily follow our intention of forgetting "old" information (it might even eliminate "new" information, because the orthogonalized rows are mostly independent of their pre-orthogonalization numbering, an approach for which a reasonable justification is not evident). Instead let us consider the following "compression" approach, based on the optimization problem

$$\begin{aligned} W \in & \underset{\tilde{W}\in\mathbb{R}^{p\times p}, \tilde{W}^* = \tilde{W}}{\operatorname{argmin}} \quad \tfrac{1}{2}||\mathcal{V}_- - \mathcal{P}^*\tilde{W}\mathcal{P}||_{HS}^2 \\ = & \underset{\tilde{W}\in\mathbb{R}^{p\times p}, \tilde{W}^* = \tilde{W}}{\operatorname{argmin}} \quad \tfrac{1}{2}||\mathcal{P}_-^* W_- \mathcal{P}_- - \mathcal{P}^*\tilde{W}\mathcal{P}||_{HS}^2 \end{aligned} \tag{308}$$

which we can solve explicitly (because it is essentially convex and quadratic). For example the least squares minimum norm solution is obviously given as

$$W_{min} := (\mathcal{P}^+)^* \mathcal{V}_-(\mathcal{P}^+) = (\mathcal{P}^+)^* \mathcal{P}_-^* W_- \mathcal{P}_- \mathcal{P}^+ \tag{309}$$

where $\mathcal{P}^+ = \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^+$ denotes the Moore-Penrose pseudoinverse of $\mathcal{P}$. For $\mathcal{P}_- = \begin{pmatrix} \mathcal{P}_{old} \\ \mathcal{P} \end{pmatrix}$ we can also use the following slightly cheaper formula

$$W := \begin{pmatrix} (\mathcal{P}\mathcal{P}^*)^+ \mathcal{P}\mathcal{P}_{old}^* & \mathcal{I} \end{pmatrix} W_- \begin{pmatrix} \mathcal{P}_{old}\mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^+ \\ \mathcal{I} \end{pmatrix} \tag{310}$$

which defines also a solution of (308), since

$$\mathcal{P}^* W \mathcal{P} = \mathcal{P}^* W_{min} \mathcal{P} \tag{311}$$

is satisfied due to the the equality

$$\begin{aligned} &\mathcal{P}^* \begin{pmatrix} (\mathcal{P}\mathcal{P}^*)^+ \mathcal{P}\mathcal{P}_{old}^* & \mathcal{I} \end{pmatrix} = \begin{pmatrix} \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^+ \mathcal{P}\mathcal{P}_{old}^* & \mathcal{P}^* \end{pmatrix} \\ =&\begin{pmatrix} \mathcal{P}^*(\mathcal{P}^+)^* \mathcal{P}_{old}^* & \mathcal{P}^*(\mathcal{P}^*)^+ \mathcal{P}^* \end{pmatrix} = \mathcal{P}^* \begin{pmatrix} (\mathcal{P}^+)^* \mathcal{P}_{old}^* & (\mathcal{P}^*)^+ \mathcal{P}^* \end{pmatrix} \\ =&\mathcal{P}^*(\mathcal{P}^+)^* \mathcal{P}_-^* \end{aligned} \tag{312}$$

which follows from the definition of the Moore-Penrose pseudoinverse. One benefit of this approach is that setting our new approximation to

$$\mathcal{V} := \mathcal{P}^* W \mathcal{P} \tag{313}$$

respects the spectrum in a certain way, as we shall see below. Even more intriguing are the low computational costs whenever $2p$ is small and we have already computed the quantity $\Phi_- := \mathcal{P}_- \mathcal{P}_-^*$ (which we would recursively update anyway, as discussed in the prior subsection.). Let us now specify what we mean by "respecting" the spectrum and state the following proposition.

Proposition 5.22. *Assume that there exist constants $\mu \leq \gamma \leq L$ with*

$$\mu \mathcal{I} \preceq \gamma \mathcal{I} + \mathcal{V}_- \preceq L \mathcal{I} \tag{314}$$

for $\mathcal{V}_- = \mathcal{P}_- W_- \mathcal{P}_-$ and $\mathcal{P}_- = \begin{pmatrix} \mathcal{P}_{old} \\ \mathcal{P} \end{pmatrix}$, then W defined as in (310) is a solution of the optimization problem (308). The semidefinite inequalities

$$\mu \mathcal{I} \preceq \gamma \mathcal{I} + \mathcal{V} \preceq L \mathcal{I} \tag{315}$$

are satisfied for the linear operator $\mathcal{V} := \mathcal{P}^* W \mathcal{P}$.

Remark 5.23. *Again Proposition 5.22 states a slightly more general result with unconstrained convex optimization in mind (For this thesis, only the case $\mu = -1, \gamma = 0$ and $L = 1$ is needed).*

Proof. Considering

$$\begin{aligned} &\mathcal{V} - (\mu-\gamma)\mathcal{I} = \mathcal{P}^*(\mathcal{P}^+)^*\mathcal{V}_-\mathcal{P}^+\mathcal{P} - (\mu-\gamma)\mathcal{I} \\ =& \underbrace{\mathcal{P}^*(\mathcal{P}^+)^*(\mathcal{V}_- - (\mu-\gamma)\mathcal{I})\mathcal{P}^+\mathcal{P}}_{\succeq 0} - (\mu-\gamma)(\mathcal{I} - \underbrace{\mathcal{P}^*(\mathcal{P}^+)^*\mathcal{P}^+\mathcal{P}}_{=\mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^+\mathcal{P})} \\ \succeq& \underbrace{(\gamma-\mu)}_{\geq 0}\underbrace{(\mathcal{I} - \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^+\mathcal{P})}_{\succeq 0} \succeq 0 \end{aligned} \tag{316}$$

yields the first semidefinite inequality by rearranging the terms. Analogously we obtain

$$(L-\gamma)\mathcal{I} - \mathcal{V} = \underbrace{(L-\gamma)}_{\geq 0}\underbrace{(\mathcal{I} - \mathcal{P}^*(\mathcal{P}\mathcal{P}^*)^+\mathcal{P})}_{\succeq 0} + \underbrace{\mathcal{P}^*(\mathcal{P}^+)^*((L-\gamma)\mathcal{I} - \mathcal{V}_-)\mathcal{P}^+\mathcal{P}}_{\succeq 0} \succeq 0 \tag{317}$$

which shows the second semidefinite inequality. ■

In practice, above approach might lead to a waste of reasonably relevant curvature information. Let us try to overcome this shortcome by noticing that we don't need to completely ignore the "old" information. In fact we can "overwrite" the old information in $\mathcal{P}_- = \begin{pmatrix} \mathcal{P}_{old} \\ \mathcal{P} \end{pmatrix}$ with the new information stored in $\mathcal{P}_+ = \begin{pmatrix} \mathcal{P}_{new} \\ \mathcal{P} \end{pmatrix}$ by considering

$$\begin{aligned} \underset{\tilde{W}=\tilde{W}^T}{\text{minimize}} \quad & ||\mathcal{P}_+^*\tilde{W}\mathcal{P}_+ - \mathcal{P}_-^*W_-\mathcal{P}_-||_{HS} \\ \text{subject to} \quad & \mathcal{P}_+^*\tilde{W}\mathcal{P}_+[S] = G \\ & -\mathcal{I} \preceq \mathcal{P}_+^*\tilde{W}\mathcal{P}_+ \preceq \mathcal{I} \end{aligned} \tag{318}$$

which can again be efficiently solved by interior point methods, as it is in fact equivalent to a (small dimensional) convex quadratic SDP whenever it is feasible (which we can always enforce by adding the "new" information appropriately). Let us give a sketch here and assume, again for simplicity, that $\mathcal{P}_+\mathcal{P}_+^*$ is invertible. By using the transformation $\xi = (\mathcal{P}_+\mathcal{P}_+^*)^{\frac{1}{2}}\tilde{W}(\mathcal{P}_+\mathcal{P}_+^*)^{\frac{1}{2}}$ and together with the definitions $\tilde{B} := (\mathcal{P}_+\mathcal{P}_+^*)^{-\frac{1}{2}}\mathcal{P}_+[S]$ $g := (\mathcal{P}_+\mathcal{P}_+^*)^{\frac{1}{2}}\mathcal{P}_+[G]$, and $B_- := (\mathcal{P}_+\mathcal{P}_+^*)^{-\frac{1}{2}}\mathcal{P}_+\mathcal{P}_-^*W_-\mathcal{P}_-\mathcal{P}_+^T(\mathcal{P}_+\mathcal{P}_+^*)^{-\frac{1}{2}}$ we can transform (318) and see that if B_+ is an optimal solution of

$$\begin{aligned} \underset{\tilde{B}=\tilde{B}^T}{\text{minimize}} \quad & ||B - B_-||_F \\ \text{subject to} \quad & Bs = g \\ & -I \preceq \tilde{B} \preceq I \end{aligned} \tag{319}$$

then the reverse transformed $W_+ := (\mathcal{P}_+\mathcal{P}_+^*)^{-\frac{1}{2}}B_+(\mathcal{P}_+\mathcal{P}_+^*)^{-\frac{1}{2}}$ will be a solution of (318). Note that if $\mathcal{V}_- = \mathcal{P}_-^*W_-\mathcal{P}_-$ satisfies $||\mathcal{V}_-|| \leq 1$ then B_- will have eigenvalues in the interval

$[-1, 1]$. We are not going to investigate this here, but there might be hints at the existence of a (not fully, but relatively explicit) low rank update formula. In our implementation we will instead use iterative methods to tackle (319) and then reconstruct an approximate solution of (318). Now that we have all ingredients together, let us focus on computations once more.

6 Numerical Results

The purpose of this section is to show that the KM iteration can, despite its (horrible and tight, see section 4) worst-case complexity, in practice often yield an acceptable convergence behavior and that this behavior may be improved by means of norm constrained quasi Newton methods. The Euclidean spaces and convex cones considered in this section, have already been discussed in section three. Let us first briefly explain the implementation and consider some numerical results afterwards. Let us denote with r the (maximal) memory size parameter (i.e. the upper bound on the rank of $\mathcal{P}^*$ is $2r$) and shorten our notation: We will use NCQNM(r) instead of norm constrained limited memory quasi Newton method with memory size r. The implementation to be considered in this subsection follows our conceptual algorithm 2.24. For $r = 0$ and constant step length $t_k \equiv t$, NCQNM(0) recovers the (inexact) KM-iteration from section four. For $r \in \mathbb{N}$ the NCQNM(r) implementation follows in principle the following path: We set the first[16] non-zero linear operator, say $\mathcal{V}^{(0)}$ such that it has the conic projections $(\Pi_{\mathcal{K}}(Z^{(0)}), \Pi_{\mathcal{K}^P}(Z^{(0)}))$ of the first iterate $(Z^{(0)})$ as eigenvectors (with eigenvalues one and minus one) and a rank lower or equal than two. After taking a step, we obtain the next linear operator $\mathcal{V}^{(k+1)}$, until the memory specified by r is exhausted, by (approximately) projecting the prior operator $\mathcal{V}_- = V^{(k)}$ first onto the next operator that satisfies the secant equation and is bounded in operator norm by one (i.e. solve (268) for $\mathcal{M} = I$). The result of this projection, say $\mathcal{V}$ is then again projected such that it has the conic projections $\Pi_{\mathcal{K}}(Z^{(k+1)}), \Pi_{\mathcal{K}^P}(Z^{(k+1)})$ of the current iterate $Z^{(k+1)}$ as eigenvectors (i.e. exploit an adapted formula of (288)). This result then defines the next $\mathcal{V}^{(k+1)}$. By using a storage scheme that only saves the last r conic projections (in $\mathcal{P}$) and updates some auxiliary matrices (W, Φ, Ψ, Θ), any explicit formation of $\mathcal{V}^{(k)}$ or $\mathcal{V}^{(k+1)}$ is successfully avoided and $||\mathcal{V}^{(k+1)}|| = 1$ can be guaranteed for all k. Afterwards we can proceed by computing the next step via recursive inversion discussed in section 5.2.3 (with some modifications for non-invertible approximations). Once the memory is exhausted, the first projection strategy is adapted such that the oldest conic projections $(\Pi_{\mathcal{K}}(Z^{(k-r)}), \Pi_{\mathcal{K}^P}(Z^{(k-r)}))$ are overwritten according to (318). Note that all subproblems involving the secant equation are only approximately solved by using Dykstra's projection algorithm [10]. For small $r \in \mathbb{N}$ the reduced overhead outweighs the benefits of a more accurate interior point method and therefore justifies its application. As in our first implementation considered at the beginning of section four, we again rescale the constant parts b and C in the definition ∇f internally, essentially preassigning the goal of finding a point $Z^{final} \in E$ such that

$$\sqrt{\frac{\left\|\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathcal{A}(\Pi_{\mathcal{K}}(Z^{final}))-b]\right\|^2}{2\|\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[b]\|^2} + \frac{\left\|(I-\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})[\Pi_{\mathcal{K}^P}(Z^{final})+C]\right\|^2}{2\|(I-\mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A})[C]\|^2}} \leq tol \tag{320}$$

[16] By default this happens after a small number (default five) of (inexact) KM iteration steps.

is satisfied for some chosen tolerance $tol > 0$, whenever the denominators are sufficiently positive. The final MATLAB implementation (1.0) is in fact quite different from the first implementation. Mathematically it is of course still a version of the conceptual algorithm 2.24. One of these differences is connected to data representation: While we used explicit sparse matrices in section four to represent the linear operators $\mathcal{A}$, $\mathcal{A}^*$, it is for many problems much more efficient to exploit the data structure and "hard code" the application of these operators, as we briefly discussed in section 2.2. We will start complementing the results from section four for sparse random SDPs. This (short) subsection will give us an expectation and we can then explain the implementation in more detail on a class of examples where above strategy makes a big difference, namely doubly non-negative relaxations for the max-stable set problem. For all computational results below we used a desktop PC from 2013 running Ubuntu 16.04 equipped with one Intel i7-4770 CPU which has four cores clocked at 3.4GHz. The MATLAB version used on this PC was 2017a. In all computations the initial point was chosen as the zero element $Z^{(0)} = 0$.

6.1 Sparse Random SDPs

We have already seen some results on the second class of problems in section 4.1, namely sparse random SDPs. Here we shall complement those tests with our final implementation (1.0), again on the test set considered in [69] which are available on Franz Rendls Webpage [74]. The difference is, apart from an upgraded PC (see above for details) as well as an upgraded operating system and MATLAB version, mainly the possibility of (cheaply) exploring the convergence behavior of NCQNM(r) for different values of $r \in \mathbb{N}$.

In fact, despite relatively large average dimensions, all 24 problems can be solved up to a tolerance $tol = 10^{-6}$ in about 3 minutes on the earlier mentioned Desktop PC, on which all test presented in this section were performed in order to maintain comparability, making it easy to generate results for different parameters and perfect for testing. We start by establishing a baseline with the KM iteration (i.e. NCQNM(0)) and then compare NCQNM(r) to that baseline with the same constant step length $t = 0.95$ and tolerance $tol = 10^{-6}$. By again employing the update scheme for $|.|_{\mathbb{S}^n_+}$ as well as the CG method for solving the linear systems (both have been explained in section four), some minor changes and additionally exploiting symmetry of the constraint matrices, the numerical results did improve significantly (in terms of accuracy and time spent.) over the earlier results to be found in table (1) . Now the real surprise is that, despite a missing line-search, NCQNM(r) reached the

	value		value
min(n)	45150	min(m)	10000
rd(median(n))	180300	rd(median(m))	50000
rd(mean(n))	212808	rd(mean(m))	59583
max(n)	500500	max(m)	150000

Table 7: Minimum, maximum and rounded average number of variables (n) and constraints(m) for sparse random SDPs considered in [69].

	$\frac{1}{2}(n_s^2 + n_s)$	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
R3002030-p3.mat	45150	20000	100	45	7.61350e+02	7.61352e+02	1.1e-06	2.8e-09	1.1e-06	1.4	5.1
R3002530-p3.mat	45150	25000	300	143	7.38378e+01	7.38384e+01	1.4e-06	7.3e-10	3.8e-06	9.3	20.7
R3001040-p4.mat	45150	10000	166	64	1.65975e+02	1.65975e+02	2.4e-07	7.5e-09	2.4e-08	4.5	10.1
R4003030-p3.mat	80200	30000	109	40	1.07214e+03	1.07214e+03	2.6e-08	6.7e-09	-1.4e-08	1.8	6.3
R4004030-p3.mat	80200	40000	167	58	8.05769e+02	8.05769e+02	1.3e-06	1.1e-09	-5.7e-07	6.6	15.6
R4001540-p4.mat	80200	15000	209	45	-6.55000e+02	-6.55000e+02	2.0e-07	7.5e-09	6.3e-08	6.2	16.9
R5003030-p3.mat	125250	30000	158	57	1.10763e+03	1.10763e+03	2.5e-07	8.2e-09	3.9e-08	3.0	10.6
R5004030-p3.mat	125250	40000	116	37	8.16611e+02	8.16611e+02	9.3e-08	6.5e-09	1.7e-08	2.5	8.9
R5005030-p3.mat	125250	50000	103	26	3.64945e+02	3.64945e+02	6.1e-08	6.1e-09	1.5e-07	2.8	9.4
R5002040-p4.mat	125250	20000	274	116	3.28004e+02	3.28004e+02	1.6e-07	7.3e-09	-1.6e-08	7.4	22.5
R6004030-p3.mat	180300	40000	171	78	3.06617e+02	3.06617e+02	2.3e-07	7.8e-09	7.5e-08	4.4	14.9
R6005030-p3.mat	180300	50000	131	44	-3.86414e+02	-3.86413e+02	2.9e-07	6.6e-09	6.6e-07	3.9	13.2
R6006030-p3.mat	180300	60000	114	37	6.41737e+02	6.41737e+02	5.6e-08	6.1e-09	6.1e-08	3.7	12.4
R6002040-p4.mat	180300	20000	392	160	1.04527e+03	1.04527e+03	1.2e-07	8.4e-09	1.8e-07	9.8	33.4
R7005030-p3.mat	245350	50000	197	7	3.13202e+02	3.13203e+02	2.1e-07	7.5e-09	1.1e-06	8.5	29.1
R7007030-p3.mat	245350	70000	122	40	-3.69558e+02	-3.69559e+02	2.6e-07	6.0e-09	-2.3e-07	5.2	17.0
R7009030-p3.mat	245350	90000	110	36	-2.67555e+01	-2.67555e+01	2.1e-08	5.4e-09	1.5e-07	5.5	17.2
R8007030-p3.mat	320400	70000	177	56	2.33140e+03	2.33140e+03	2.3e-07	6.8e-09	7.0e-08	9.5	31.3
R80010030-p3.mat	320400	100000	116	8	2.25929e+03	2.25929e+03	1.2e-07	5.5e-09	2.7e-08	7.8	25.8
R80011030-p3.mat	320400	110000	113	16	1.85792e+03	1.85792e+03	2.6e-08	5.3e-09	6.3e-09	7.9	25.2
R90010030-p3.mat	405450	100000	150	15	9.54223e+02	9.54223e+02	2.6e-07	5.8e-09	2.6e-07	12.2	40.6
R90014030-p3.mat	405450	140000	113	15	2.31983e+03	2.31983e+03	3.7e-08	4.7e-09	-1.3e-08	10.6	33.0
R100010030-p3.mat	500500	100000	204	55	3.09636e+03	3.09636e+03	2.0e-07	6.5e-09	3.8e-08	19.0	61.6
R100015030-p3.mat	500500	150000	119	26	1.05289e+03	1.05289e+03	2.2e-07	4.8e-09	7.0e-08	12.9	41.2

Table 8: Final implementation (1.0) performance with $r = 0$, $t = 0.95$, $tol = 10^{-6}$ and $Z^{(0)} = 0$ on random sparse SDPs considered in [56]. Results differ from (1) due to improved implementation and/or rounding errors. Wall-clock (**time**) and total CPU-time (**cpu**) is given in seconds.

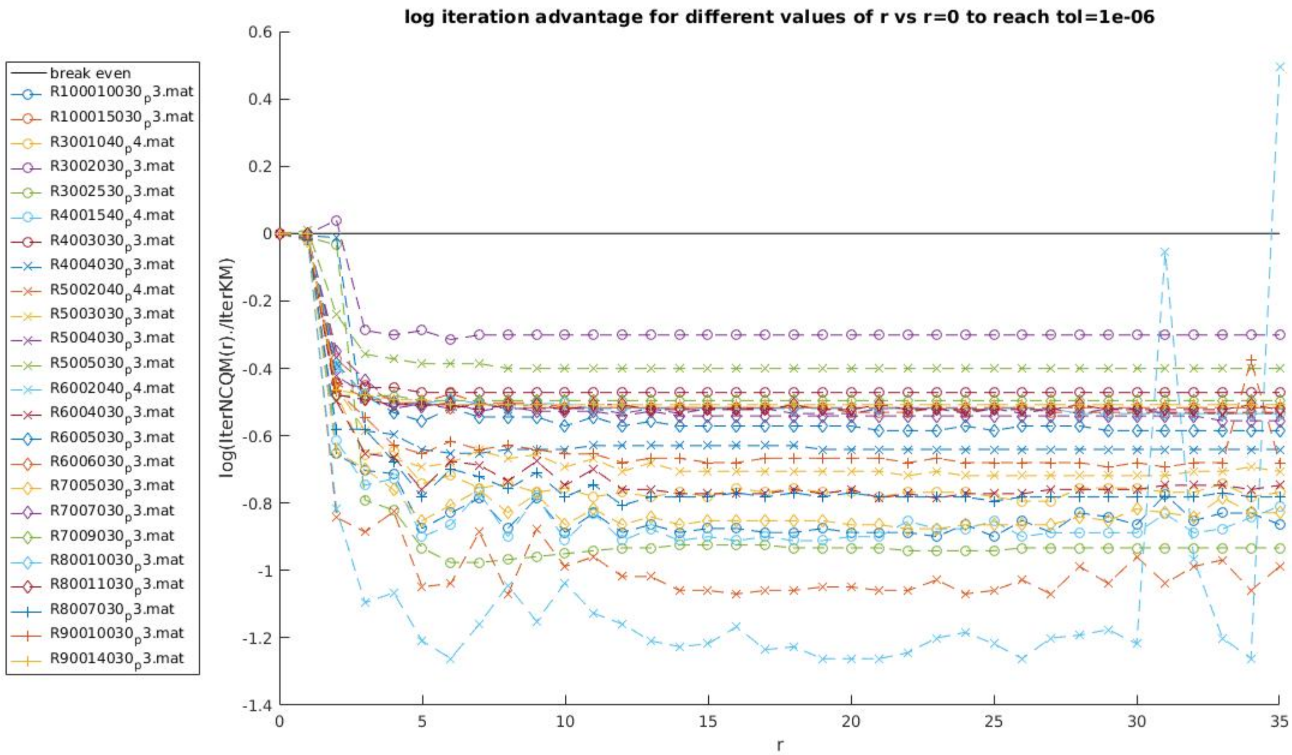


Figure 9: MATLAB plot comparing the (natural) logarithmic advantage (in terms of number of iterations) of NCQNM(r) for different values of r vs. NCQNM(0) constant step length $t = 0.95$ on sparse random SDPs. Every entry below the black line implies that NCQNM(r) needed less iterations than NCQNM(0) to reach a tolerance of $tol = 10^{-6}$. The minimal average number of iteration to solve all considered problems with NCQNM(r), is attained for $r = 22$ leading to a speed up, with respect to the number of iterations, of about 210%.

desired tolerance for all problems and for all values of $r \in \{0, ..., 35\}$, which is an interesting improvement over regular quasi Newton methods behavior and may be due to the norm constraint. However for larger memory size $r \in \mathbb{N}$ the iteration may become less stable due to multiple reasons, some of which will be addressed below. In figure (9) we compare the relative number of iterations it takes to reach the desired tolerance $tol = 10^{-6}$. Note that the value $r = 1$ implies that we are only using the Eigenupdate (288) and do not benefit from the secant equation, which results in additional time spent but not in a convergence improvement. However for $r \in \{2, 3, 4, 5, 6\}$ we do see significantly lower number of iterations to reach convergence. We can also compare the relative time it takes to reach the desired tolerance of $tol = 10^{-6}$. In figure (10) we can clearly see the up- and downsides of larger values of r. Choosing large values of $r \in \mathbb{N}$ does not necessarily improve convergence rates significantly (here the improvement seems negligible for r larger then 20), but it certainly raises the computational time per iteration. On the other hand if, for example, we average the number of iterations over all problems, then the minimal number of

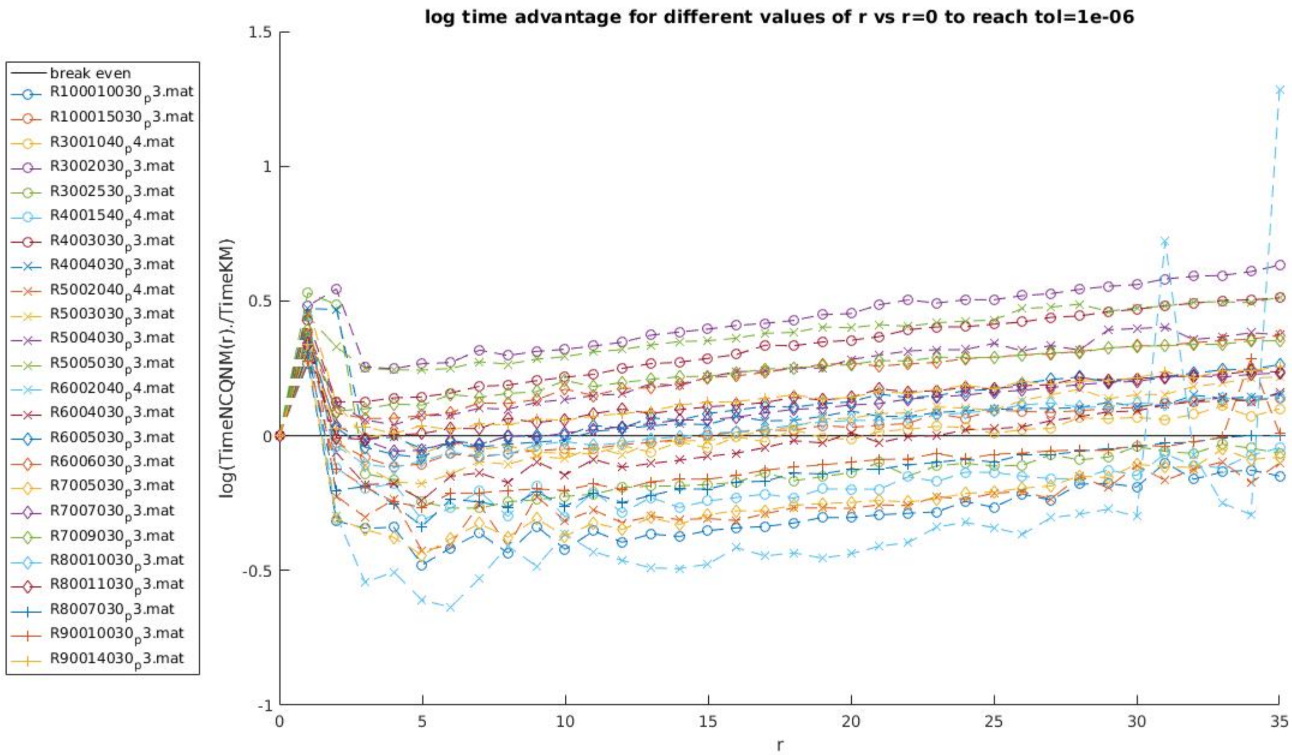


Figure 10: MATLAB plot comparing the (natural) logarithmic advantage (in terms of time) of NCQNM(r) for different values of r vs. NCQNM(0) constant step length $t = 0.95$ on sparse random SDPs. Every entry below the black line implies that NCQNM(r) needed less time than NCQNM(0) to reach a tolerance of $tol = 10^{-6}$. The minimal average time to solve all considered problems with NCQNM(r), is attained for $r = 6$ leading to a speed up in terms of time of about 20%.

iterations for all problems is actually attained for $r = 22$. While this trade off seems less beneficial for small dimensions, it might, for large dimensions, be quite tempting to choose r as large as the computer memory allows. Note however that larger r do not only result in larger subproblems, but also in less adaptiveness: Old curvature information are kept longer in memory although they might not be relevant anymore or even wrong due to the non-differentiability of $|.|_{\mathcal{K}}$. Now as mentioned earlier let us next try to explain some of the implementation ideas and its usage for another set of test problems.

6.2 Maximum Stable Set Relaxations

In this subsection we consider numerical results for computing the Lovasz numbers $\theta(G)$, $\theta(\bar{G})$ and Lovasz-Schrijver numbers $\theta^+(G)$, $\theta^+(\bar{G})$ for simple graphs $G = (V, E(G))$ for which we already stated the problem formulation in section 3.2.2. Our test graphs consist of 62 graphs taken from the second DIMACS challenge [34]. Note that we are considering the original graphs (indicated by G and their complement graphs (indicated by $\bar{G}$). Their roles are sometimes interchanged in the literature (for example in [99]), which does not pose any problems as long as we are consistent.[17] At the time of writing this passage, the original source download link [36] seems to be broken, but an alternative is given by [37]. Since our problem formulation (1) is posed as a minimization problem, but the max stable set relaxations (104) and (105) are posed as maximization problems we have to switch signs and define $C := -J$. Recall that the scaling used in section 3.2.1 results in $\mathcal{A}\mathcal{A}^*$ being equal to some identity operator of appropriate size, implying that there is no additional computational cost in its inversion. Note however, that the applied scaling affects the relative primal and relative dual infeasibility measures

$$R_D := \frac{||C + S^{(*)} - \mathcal{A}(y^{(*)})||_E}{1 + ||C||_E}, \quad R_P := \frac{||b - \mathcal{A}(X^{(*)})||_2}{1 + ||b||_2}, \quad gap := \frac{b^T y^{(*)} - C \bullet X^{(*)}}{1 + |b^T y^{(*)}| + |C \bullet X^{(*)}|} \tag{321}$$

where $X^{(*)}, S^{(*)} := -Y^{(*)}$ and $y^{(*)}$ are computed from the final iterate Z^{final} according to corollary 2.9. We will therefore still report these quantities, but avoid using them in our analysis. The easiest way of presenting the implementation might be an "applied" example: After starting MATLAB and switching to the solvers main directory, the following code will load the data, create the Lovasz problem, and approximate it with default options via the (inexact) KM iteration:

```
pathsetup; % Sets all required paths
Libnum=3;  % Corresponds to dimacs2nd
Probnum=2; % Second dataset, here brock200-2.clq
AG=LoadTestProblem(Libnum,Probnum); % Load data, adjacency matrix AG
[Afunc,b,c,K]=CellLovasz(AG);       % Create Lovasz problem
[Var,info]=main(Afunc,b,c,K);       % Solve problem with default options
```

Here, `Afunc` is a MATLAB function handle that can, depending on its input, evaluate the operators action $\mathcal{A}(X)$, $\mathcal{A}^*(y)$ or $(\mathcal{A}\mathcal{A}^*)^{-1}y$ for the Lovasz number problem (105). For the default options the step length is fixed to $t = 0.95$, the maximal number of iterations to $maxit = 10^6$ and the tolerance $tol = 10^{-6}$. This can be adjusted (as well as many other options) by additionally using a separate options input:

```
options=setoptions(struct); % Load default options
options.maxit=3000;         % Adjust maximal number of iterations
options.tol=0.0001;         % Adjust tolerance
options.stepsize=0.5;       % Adjust step length t
options.maxtime=600;        % Set maximal time to 10 minutes
[Var,info]=main(Afunc,b,c,K,options); % Solve prob with adjusted options
```

In both cases, the main program prints some information and (upon normal termination) returns the struct `Var` in which all final iterates are stored in cells, as well as the struct `info` in which additional information is provided. Note that only the lower triangular part

[17] Let us note (as an anchor), that the original c-fat200-1 graph has 1534 edges.

of symmetric matrix variables are stored in a column vector. The execution time should be less than half a second on a reasonably new computer running Linux and MATLAB 2016b or 2017a (Probably also on other systems and versions, but there was no testing). To activate the usage of the norm constrained quasi Newton method, we can use

```
options=setoptions(struct); % Load default options
options.qnewt.usequnewt=1;  % Activate NCQNM
options.qnewt.maxr=10;      % Set max approximation rank to ten
[Var,info]=main(Afunc,b,c,K,options); % Solve prob with adjusted options
```

Note however that there is absolutely no theoretical convergence guarantee (which is the reason for turning this option off by default) and that there is no line-search implemented (which would be rather expensive computationally) , i.e. the default step length ($t = 0.95$) is used (here). Interestingly the method also converges reasonably well on this second set of examples nevertheless. Instead of considering large tables with results here, let us try to give an idea first about the test set in terms of dimensions (n) as well as number of constraints (m) and secondly about convergence behavior in terms of both number of iterations (iter) and timing (sec) for different values of r. For completeness the mentioned large tables (created with a modified version of matrix2latex.m [45]) are included in the Appendix. The DIMACS graphs vary heavily in the number of vertices and edges, rendering the resulting Lovasz problems (105) and Lovasz-Schrijver problems (104) a great test set. For each problem the Euclidean space's dimension (n) is equal to the number of (non-redundant) variables of the semidefinite block ($\frac{n_s(n_s+1)}{2}$, note that n_s is equal to the number of vertices) plus the number of variables of the nonnegative block (n_l) (zero for Lovasz problems, equal to the number of edges in the complementary graph considered for the Lovasz-Schrijver problems). The number of constraints is determined by one plus the number of edges plus the number variables in the non-negative block. In order to give a rough idea, minimum, maximum, rounded mean and rounded median of the dimensions are reported in the table below as well as two histograms, the exact values for each individual problem can be found in the Appendix.

Now when it comes to analyzing the convergence behavior, let us start again by establishing a baseline with the KM iteration (NCQNM(0)) for a tolerance of $tol = 10^{-6}$. What becomes very clear from the large differences in mean and median values is that there are a couple of problems for which the KM iteration needs a large amount of iterations respectively time to reach the required tolerance. This is not an effect of rounding errors, but of highly degenerated problems resulting from the p_hat500-2, p_hat700-2, p_hat1000-2 and p_hat1500-2 graphs. For these problems it seems that (many) eigenvalues of the iterates SDP block ($Z^{(k)}$) are clustered around zero, resulting especially in non-differentiability of the generalized absolute value at those points and probably also at the optimal solution. This implies that other methods, for example interior point methods, are also likely to suffer from slow convergence for these problems. Note however that other methods do usually not rely on the convergence criterion (320) and therefore may return early with an accuracy that users might accept as sufficient, but that might be strictly worse than the goal defined by (320). Note that in all examples the convergence rates are far better than the theoretical worst-case of the KM-iteration ($O(\frac{1}{\sqrt{k}})$), proven in section 4.1, predicts. In fact the convergence rate

	$\theta(G)$	$\theta^+(G)$	$\theta(\bar{G})$	$\theta^+(\bar{G})$
min(n)	406	574	406	616
rd(median(n))	80200	104136	80200	136060
rd(mean(n))	270834	36813	270834	448154
max(n)	5649841	6676423	5649841	10269739
min(m)	211	379	169	379
rd(median(m))	43307	79801	23428	79801
rd(mean(m))	177322	270301	92980	270301
max(m)	4619899	5646481	1026583	5646481

Table 9: Minimum, maximum and rounded average number of variables (n) and constraints(m) for Lovasz problems and Lovasz-Schriver problems of 62 DIMACS graphs as well as their complement graphs.

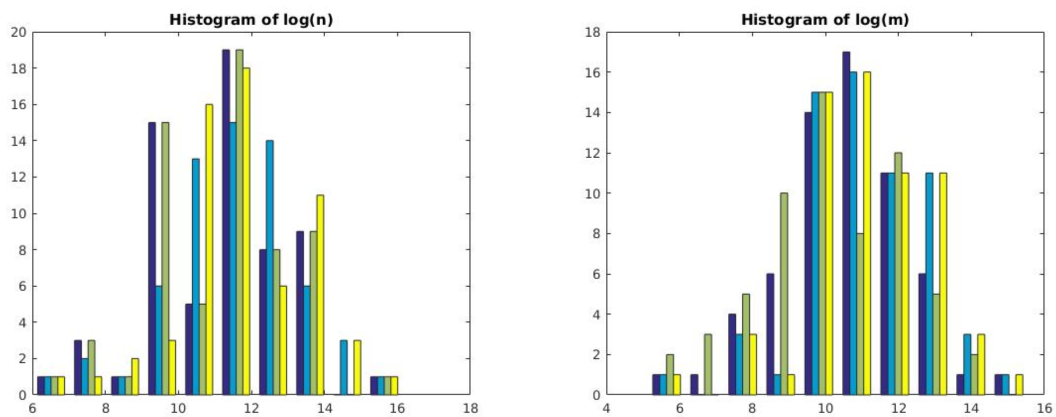


Figure 11: MATLAB histograms (`hist`) of (natural) logarithmic dimension n and number of constraints m for Lovasz problem of G (dark blue), Lovasz-Schriver problem of G (blue), Lovasz problem of $\bar{G}$ (green) and Lovasz-Schriver problem of $\bar{G}$ (yellow).

is even better than of the bound on the Halpern iteration $O(\frac{1}{k})$ considered in section 4.2.1. One may certainly argue that the test set is not sufficiently large to experience worst-case behavior, but another possible interpretation is that worst-case examples possess a certain kind of rareness (as suggested by section 4.2.2).

	$\theta(G)$	$\theta^+(G)$	$\theta(\bar{G})$	$\theta^+(\bar{G})$
min(iter)	99	135	100	131
rd(median(iter))	1463	1942	2303	4096
rd(mean(Iter))	7295	9750	17135	14917
max(Iter)	65543	96758	301413	90934
min(sec)	0.1699	0.2970	0.0945	0.1320
rd(median(sec))	6	8	24	36
rd(mean(sec))	658	903	557	782
max(sec)	18423	24939	19729	27150

Table 10: Rounded average number of iterations (iter) and seconds (sec) to reach $tol = 10^{-6}$ via the KM-iteration for 62 dimacs graphs. The high differences between median and mean values are mainly causes by the degenerated problems p_hat500-2.clq, p_hat700-2.clq, p_hat1000-2.clq and p_hat1500-2.clq .

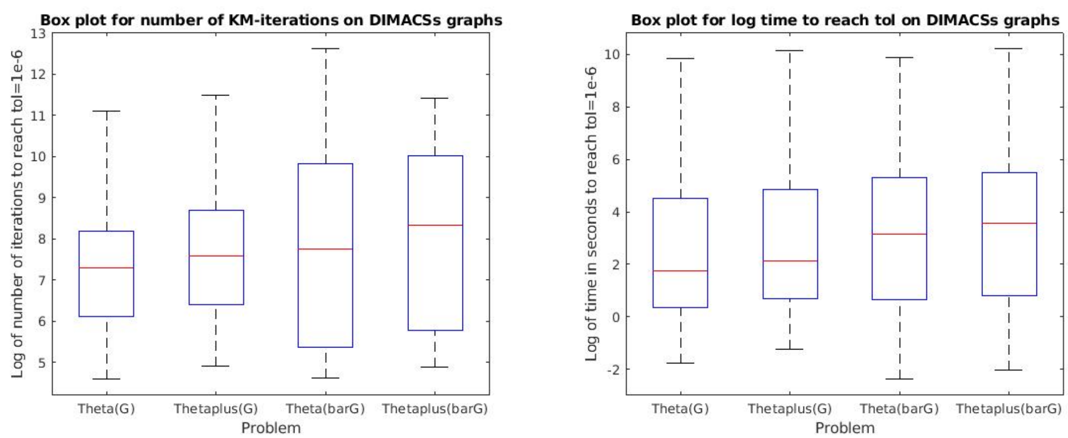


Figure 12: MATLAB boxplots (`boxplot`) of (natural) logarithmic number of iterations and of (natural) logarithmic time (in seconds) to reach $tol = 10^{-6}$ via (inexact) KM-iteration aka NCQNM(0). From MATLAB documentation: On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively.

Keeping in mind that the tolerance of ($tol = 10^{-6}$) might smaller than needed, as the Lovasz- respectively Lovasz-Schrijver numbers are often rounded down anyway, we might also consider increased tolerances. We do this exemplarily for the Lovasz-Schrijver-problem for all 62 graphs G in the box plot below.

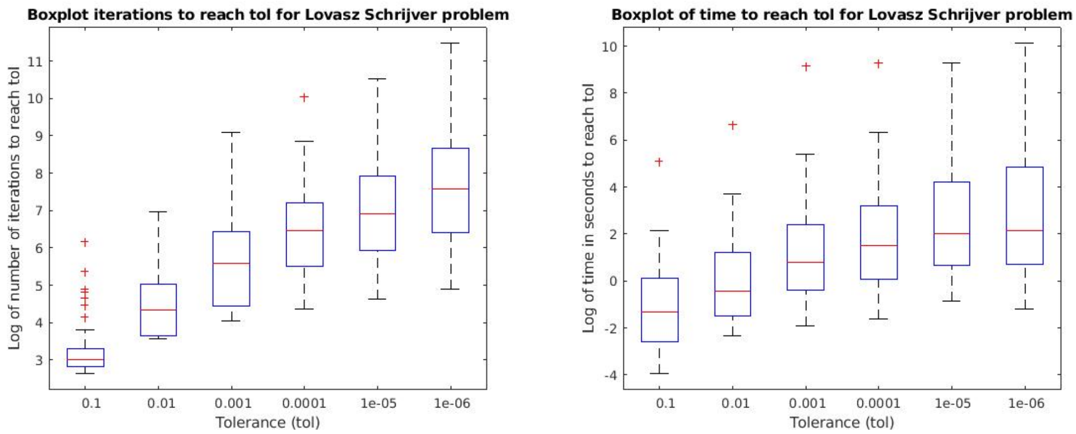


Figure 13: MATLAB boxplots (`boxplot`) of (natural) logarithmic number of iterations and of (natural) logarithmic time (in seconds) to to reach different tolerances for the Lovasz-Schrijver problems $\theta^+(G)$ of 62 DIMACS graphs via (inexact) KM-iteration aka NCQNM(0). From MATLAB documentation: On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol.

The real question to be answered now is whether for $r \in \mathbb{N}$ the NCQNM(r) can be competitive or better than the KM-iteration. While there is of course no global convergence guarantee (and any local convergence guarantee may depend on strong assumptions), the method(s) seem(s) to converge surprisingly often, even with fixed step length. For the Lovasz-Schrijver problem and values $r \in \{5, 10, 15\}$ we can visualize the results by considering the following relative results, starting with comparing the number of iterations needed to reach the desired tolerance $tol = 10^{-6}$.
Note that in general the number of iterations is (obviously) not monotone decreasing in $r \in \mathbb{N}$. Sometimes smaller values of $r \in \mathbb{N}$ may actually result in faster convergence. This may be explained by "jumps" over points of non-differentiability of ∇f, where old and inaccurate information is "forgotten" faster for smaller values of r. For the graph c-fat200 NCQNM(r) does not converge for values of $r \in \{5, 10, 15\}$ (within 100000 iterations), but further testing revealed, that it does for larger values of r (for example $r = 25$). Choosing r adaptively may possibly resolve these issues in a future version.

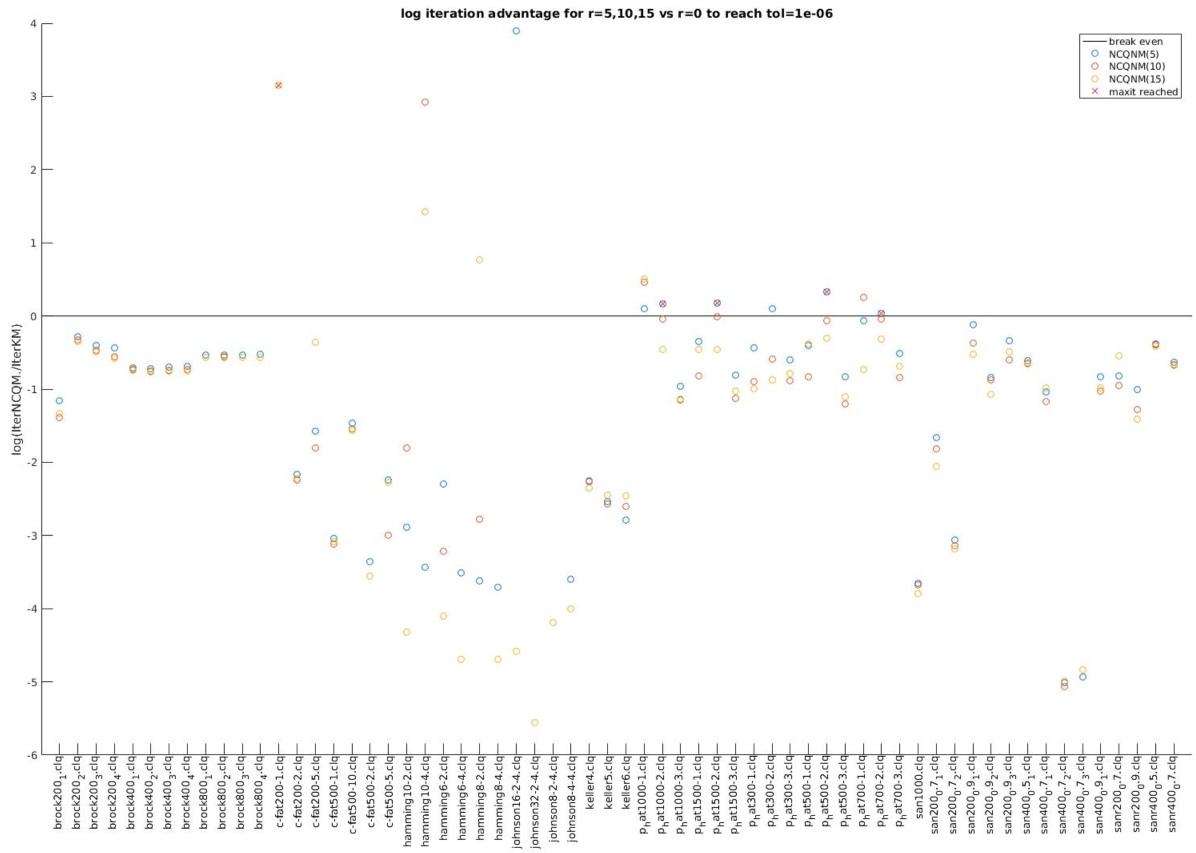


Figure 14: MATLAB plot comparing the (natural) logarithmic advantage (in terms of number of iterations) of NCQNM(r) for $r \in \{5, 10, 15\}$ vs. NCQNM(0). Every circle 'o' below the black line implies that NCQNM(r) needed less iterations than NCQNM(0) to reach a tolerance of $tol = 10^{-6}$. Note that circles may overlap. We set a maximal number of iterations of 100000 and non-convergence is marked with a red cross 'x' through the circle.

We can also compare the relative time it takes to reach the desired tolerance $tol = 10^{-6}$.

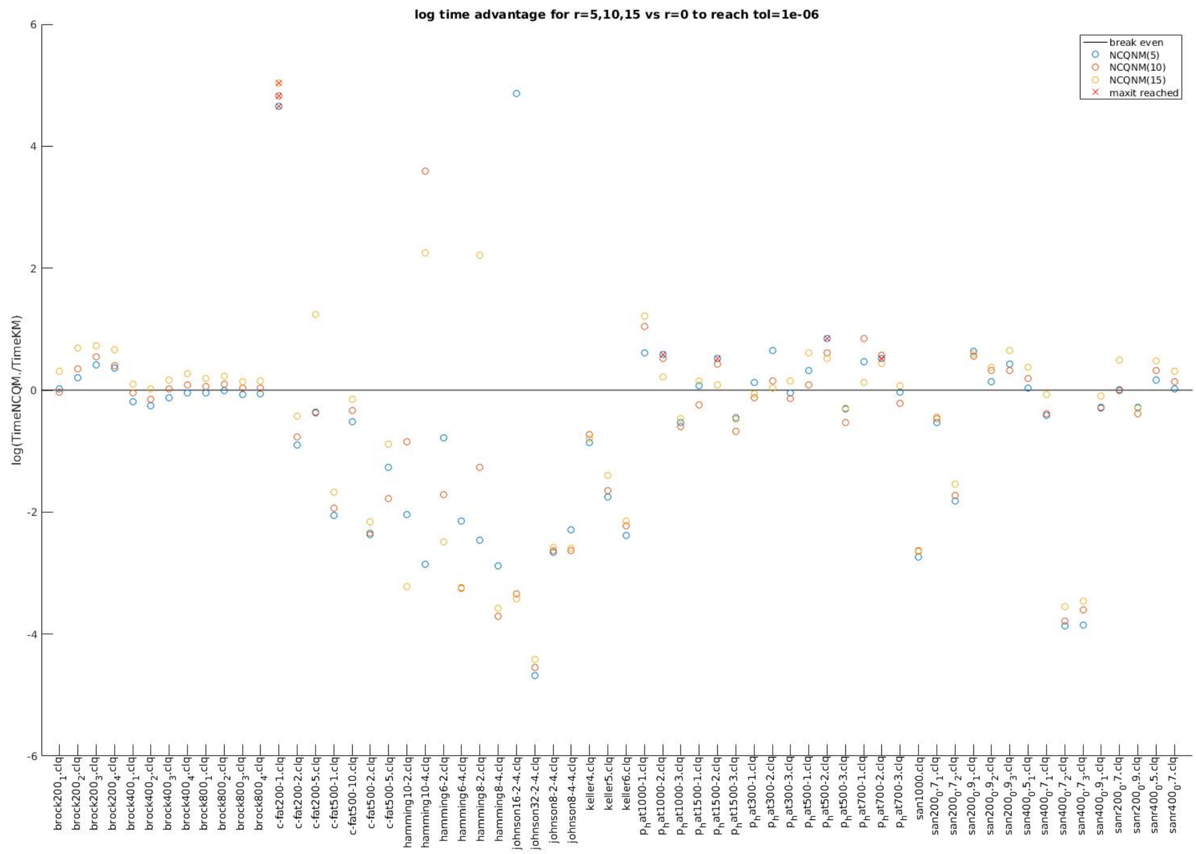


Figure 15: MATLAB plot comparing the (natural) logarithmic advantage (in terms of time) of NCQNM(r) for $r \in \{5, 10, 15\}$ vs. NCQNM(0). Every circle 'o' below the black line implies that NCQNM(r) needed less time than NCQNM(0) to reach a tolerance of $tol = 10^{-6}$. We set a maximal number of iterations of 100000 and non-convergence is marked with a red cross 'x' through the circle.

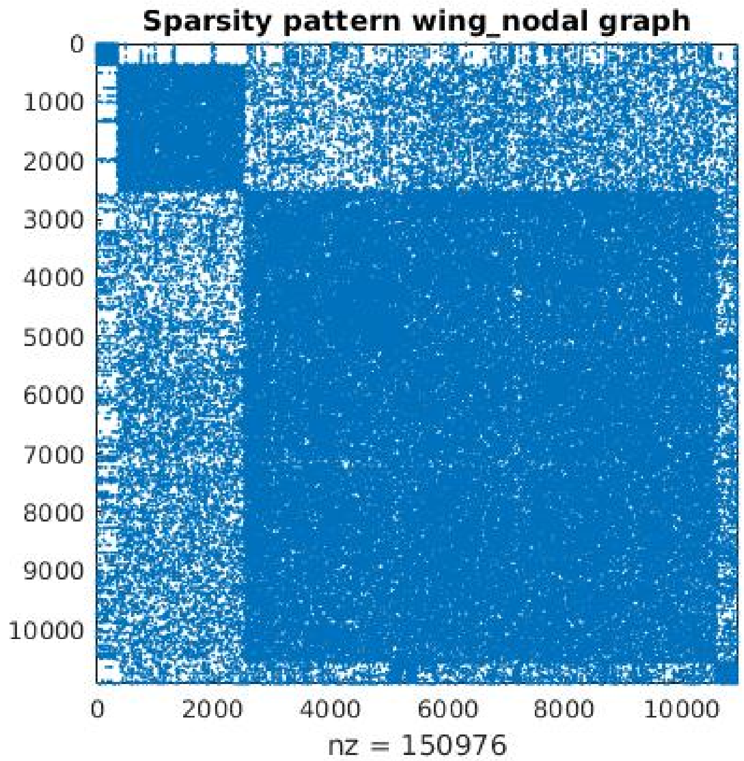


Figure 16: Sparsity pattern of (symmetric) adjacency matrix for wing_nodal graph.

Now to explore the limitations of our approach, we can also consider larger graphs, for example the undirected graph wing_nodal which has 10937 vertices and 75488 edges (this graph was downloaded from the SuiteSparse Matrix Collection [21], but was originally created by C. Walshaw for the 10th DIMACS challenge [35].) The resulting problem Lovasz-Schrijver problem (104) has roughly 112 million variables and 60 million constraints. These dimensions and the 32 Gigabyte of available memory on the desktop PC limits the size of r effectively to six (larger values may be realized by further code improvements). The time limit was set to 24 hours, which due to implementation reasons may be slightly exceeded in reality. Details on further numerical experiments are omitted here, as I believe that this

	$\frac{1}{2}(n_s^2 + n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
wing_nodal	59814453	59728028	59803517	843	0	-2.27507e+03	-2.27508e+03	1.3e-05	1.1e-03	-3.4e-07	86469.8	327891.9

Table 11: Algorithm-Performance NCQNM(6) for Lovasz-Schrijver problem on wing_nodal graph. The step length was set to $t = 0.95$ and max time was reached after about 24 hours. The residual (320) was still 1.39170e-03.

would in total water down both, the good numerical results and the (optimal) theoretical results from before. Let us therefore close this section by noting that, despite possible improvements, the results look, especially for (relatively) large dimensions, quite promising. This raises hopes that limited memory norm constrained quasi Newton methods may gain recognition and popularity in the future.

7 Conclusion

In this thesis we tackled the important class of linear conic optimization problems by means of projection based algorithms. We started with laying the foundations by recalling some well known results and problems. We have then seen that there exists a gap between worst-case complexities and average-case complexities for these methods and more generally FSMs, by exemplarily calculating the worst case complexity (which was priorly unknown) for the KM iteration with constant step length. This calculation revealed parts of a broad (and new) concept, which can be used in the future to not only answer fundamental complexity questions but also help in the development of new methods. Specifically, as first order methods gain more and more popularity arising from new applications in data science, machine learning, neural networks and other fields, it might be useful for developing algorithms that enjoy good average and worst case complexities simultaneously, potentially reducing the earlier mentioned gap. The second main contribution of this thesis is the development of efficient limited memory norm constrained quasi Newton methods, that, different from FSMs, represent a member from the class of adaptive and therefore more heuristic methods. While well chosen FSMs are numerically solid as a rock (i.e. converge whenever there is a fixed/saddle point that they can converge to), adaptive methods are less stable but often much faster in practice. The final implementation (1.0) considered in the previous section seems to take one further step towards the goal of ”getting the best” or at least ”avoiding the worst” from both worlds.

Acknowledgment

While writing this passage I realize that there are countless people that I am grateful to. I will therefore not thank each and everyone personally here, as the list would be unreasonably long and clearly misplaced (note, that this might be on the Internet forever and not everyone is alright with that). However, even if I do not mention your name here, let me say that your support is appreciated and not forgotten. Exceptions can be made and therefore I would like to thank the, for this thesis, (second) most important person, my supervisor and thesis adviser Florian Jarre. Florian has made this thesis possible, by giving me a proper education in optimization, a job right after my Master’s thesis and countless opportunities along the way especially when it came to choosing research topics or making scientifically related business trips. He was always happy to support me scientifically, while still giving me a lots of freedom. Most importantly, he did not put (much) pressure on me when it came to finishing my thesis. Could I have finished earlier? Certainly yes. Would the thesis have been similar in terms of research value? Absolutely not! In times where morons think, that science is a machine that transforms work into valuable papers linearly (the two factors are of course related, but in an unknown and highly non-linear way), Florian’s characteristics are rather rare, but greatly appreciated. Secondly I would like to address the current and former members of the mathematical institute at the Heinrich-Heine university, i.e. my coworkers, which have been great to work with. Finally I like to mention that ”Computational support and infrastructure was provided by the Centre for Information and Media Technology (ZIM) at the University of Düsseldorf (Germany).”

8 References

[1] Baillion, J., & Bruck, R. E. (1996). The Rate of Asymptotic Regularity Is O($1/\sqrt{n}$). Lecture Notes in Pure and Applied Mathematics, 51-82.

[2] Bauschke, H. H. (1996). Projection algorithms and monotone operators (Doctoral dissertation, Theses (Dept. of Mathematics and Statistics)/Simon Fraser University).

[3] Burer, S. (2009). On the copositive representation of binary and continuous nonconvex quadratic programs. Mathematical Programming, 120(2), 479-495.

[4] Burer, S., & Monteiro, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. Mathematical Programming, 95(2), 329-357.

[5] Bravo, M., Cominetti, R., & Pavez-Sign, M. (2017). Rates of convergence for inexact Krasnosel'skii-Mann iterations in Banach spaces. arXiv preprint arXiv:1705.09340.

[6] Browder, F. E. (1966). Existence and approximation of solutions of nonlinear variational inequalities. Proceedings of the National Academy of Sciences, 56(4), 1080-1086.

[7] Browder, F. E. (1967). Convergence of approximants to fixed points of nonexpansive nonlinear mappings in Banach spaces. Archive for Rational Mechanics and Analysis, 24(1), 82-90.

[8] Browder, F. E. (1967). Convergence theorems for sequences of nonlinear operators in Banach spaces. Mathematische Zeitschrift, 100(3), 201-225.

[9] Borwein, J. M., & Zhu, Q. J. (1999). Multifunctional and functional analytic techniques in nonsmooth analysis. In Nonlinear analysis, differential equations and control (pp. 61-157). Springer, Dordrecht.

[10] Boyle, J. P., & Dykstra, R. L. (1986). A method for finding projections onto the intersection of convex sets in Hilbert spaces. In Advances in order restricted statistical inference (pp. 28-47). Springer, New York, NY.

[11] Clarke, F. H. (1990). Optimization and nonsmooth analysis (Vol. 5). Siam.

[12] Chen, X. D., Sun, D., & Sun, J. (2003). Complementarity functions and numerical experiments on some smoothing Newton methods for second-order-cone complementarity problems. Computational Optimization and Applications, 25(1-3), 39-56.

[13] Choi, S. C. T., Paige, C. C., & Saunders, M. A. (2011). MINRES-QLP: A Krylov subspace method for indefinite or singular symmetric systems. SIAM Journal on Scientific Computing, 33(4), 1810-1836.

[14] Cominetti, R., Soto, J. A., Vaisman, J. (2014). On the rate of convergence of Krasnoselski-Mann iterations and their connection with sums of Bernoullis. Israel Journal of Mathematics, 199(2), 757-772.

[15] CPLEX, I. (2005). High-performance software for mathematical programming and optimization.

[16] Dantzig, G. B. (2002). Linear programming. Operations research, 50(1), 42-47.

[17] Dantzig, G.B. (1949). Programming in a linear structure, Econometrica 17 73-74

[18] Wood, M.K. & Dantzig, G.B. (1949). Programming of interdependent activities, I, General discussion, Econometrica 17 193-199

[19] Dantzig, G.B. (1949). Programming of interdependent activities, II, Mathematical model, Econometrica 17 200-211

[20] Davi, T. (2012). Lösung großer konischer Programme mit Hilfe primal-dualer Methoden. Diss. Universitäts-und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf

[21] Davis, T., & Hu, Y. The SuiteSparse Matrix Collection (formerly known as the University of Florida Sparse Matrix Collection), University of Florida and AT&T Research. https://sparse.tamu.edu/

[22] Dür, M. (2010). Copositive programminga survey. In Recent advances in optimization and its applications in engineering (pp. 3-20). Springer, Berlin, Heidelberg.

[23] De Klerk, E., & Pasechnik, D. V. (2002). Approximation of the stability number of a graph via copositive programming. SIAM Journal on Optimization, 12(4), 875-892.

[24] Drori, Y., & Teboulle, M. (2014). Performance of first-order methods for smooth convex minimization: a novel approach. Mathematical Programming, 145(1-2), 451-482.

[25] Freund, R. W., & Jarre, F. (2004). A sensitivity result for semidefinite programs. Operations Research Letters, 32(2), 126-132.

[26] Optimization, Gurobi. "Gurobi Optimizer 5.0." (2012).

[27] Halpern, B. (1967). Fixed points of nonexpanding maps. Bulletin of the American Mathematical Society, 73(6), 957-961.

[28] Helmberg, C., & Rendl, F. (2000). A spectral bundle method for semidefinite programming. SIAM Journal on Optimization, 10(3), 673-696.

[29] Hintermüller, M. (2010). Semismooth Newton methods and applications. Department of Mathematics, Humboldt-University of Berlin.

[30] Higham, N. J. (2015). The Singular Value Decomposition.

[31] Jarre, F., & Lieder, F. (2017). A Derivative-Free and Ready-to-Use NLP Solver for Matlab or Octave.

[32] Jarre, F., & Rendl, F. (2008). An augmented primal-dual method for linear conic programs. SIAM Journal on Optimization, 19(2), 808-823.

[33] Jarre, F., & Stoer, J. (2013). Optimierung. Springer-Verlag.

[34] Johnson, D. S., & Trick, M. A. (Eds.). (1996). Cliques, coloring, and satisfiability: second DIMACS implementation challenge, October 11-13, 1993 (Vol. 26). American Mathematical Soc..

[35] Bader, D. A., Meyerhenke, H., Sanders, P., & Wagner, D. (2013). Graph Partitioning and Graph Clustering: 10th DIMACS Implementation Challenge, vol. 588. American Mathematical Society, 7, 210-223.

[36] Dimacs web page: ftp://dimacs.rutgers.edu/pub/challenge/graph/benchmarks/clique/

[37] Alternative download of Dimacs graphs, Penn State Harrisburg: https://turing.cs.hbg.psu.edu/txn131/clique.html

[38] Kanzow, C., Ferenczi, I., & Fukushima, M. (2009). On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity. SIAM Journal on Optimization, 20(1), 297-320.

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

[40] Khachian, L. G. (1979). A polynomial algorithm in linear programming. Doklady Akademii Nauks SSR, 244(5), 1093-1096.

[41] Kirszbraun, M. D. (1934). Über die zusammenziehende und Lipschitzsche Transformationen". Fund. Math. 22: 77-108.

[42] Kummer, B. (1992). Newtons method based on generalized derivatives for nonsmooth functions: convergence analysis. In Advances in optimization (pp. 171-194). Springer, Berlin, Heidelberg.

[43] Krasnoselski, M.A. (1955). Two remarks on the method of successive approximations, Uspekhi Mat. Nauk 10:1(63), 123-127.

[44] Kummer, B. (1988). Newtons method for non-differentiable functions. Advances in mathematical optimization, 45, 114-125.

[45] Koehler, M (2004) matrix2latex.m available at https://de.mathworks.com/matlabcentral/fileexchange/4894-matrix2latex

[46] Kohlenbach, U. (2001). A quantitative version of a theorem due to Borwein-Reich-Shafrir, Numer. Funct. Anal. and Optimiz. 22, 641-656

[47] Kohlenbach U. (2003). Uniform asymptotic regularity for Mann iterates, J. Math. Anal. Appl. 279, 531-544.

[48] Kong, L., Tunel, L., & Xiu, N. (2009). Clarke generalized Jacobian of the projection onto symmetric cones. Set-Valued and Variational Analysis, 17(2), 135-151.

[49] Lazar, M., & Jarre, F. (2016). Calibration by optimization without using derivatives. Optimization and Engineering, 17(4), 833-860.

[50] Web page of Florian Jarre: http://www.opt.uni-duesseldorf.de/ jarre/dot/dot.html

[51] Leustean, L. (2007). Rates of Asymptotic Regularity for Halpern Iterations of Nonexpansive Mappings. J. UCS, 13(11), 1680-1691.

[52] Liang, J., Fadili, J., & Peyr, G. (2016). Convergence rates with inexact non-expansive operators. Mathematical Programming, 159(1-2), 403-434.

[53] Lieder, F. (2017). On the Convergence Rate of the Halpern-Iteration. submitted.

[54] Löfberg, J. (2004, September). YALMIP: A toolbox for modeling and optimization in MATLAB. In Computer Aided Control Systems Design, 2004 IEEE International Symposium on (pp. 284-289). IEEE.

[55] Malick, J., & Sendov, H. S. (2006). Clarke generalized Jacobian of the projection onto the cone of positive semidefinite matrices. Set-Valued Analysis, 14(3), 273-293.

[56] Malick, J., Povh, J., Rendl, F., & Wiegele, A. (2009). Regularization methods for semidefinite programming. SIAM Journal on Optimization, 20(1), 336-356.

[57] Mann W.R. (1953). Mean value methods in iteration, Proceedings of the American Mathematical Society 4(3), 506-510.

[58] Martin-Márquez, V. (2010). Fixed point approximation methods for nonexpansive mappings: optimizations problems. Universidad de Sevilla.

[59] Mifflin, R. (1977). Semismooth and semiconvex functions in constrained optimization. SIAM Journal on Control and Optimization, 15(6), 959-972.

[60] Moreau, J. J. (1962). Décomposition orthogonale dun espace hilbertien selon deux cones mutuellement polaires. CR Acad. Sci. Paris, 225, 238-240.

[61] Mosek, A. P. S. (2010). The MOSEK optimization software. Online at http://www. mosek. com, 54(2-1), 5.

[62] Nakatsukasa, Y., Bai, Z., & Gygi, F. (2010). Optimizing Halley's iteration for computing the matrix polar decomposition. SIAM Journal on Matrix Analysis and Applications, 31(5), 2700-2720.

[63] Nemirovski, A. (2004). Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM Journal on Optimization, 15(1), 229-251.

[64] Nesterov, Y. (1998). Introductory Lectures on Convex Programming : Volume I: Basic course.

[65] Nesterov, Y. (1983, February). A method of solving a convex programming problem with convergence rate O $(1/k^2)$. In Soviet Mathematics Doklady (Vol. 27, No. 2, pp. 372-376).

[66] Oren, S. S. (1974). On the selection of parameters in self scaling variable metric algorithms. Mathematical Programming, 7(1), 351-367.

[67] Ortega, J. M., & Rheinboldt, W. C. (1970). Iterative solution of nonlinear equations in several variables (Vol. 30). Siam.

[68] Paige, C. C., & Saunders, M. A. (1975). Solution of sparse indefinite systems of linear equations. SIAM journal on numerical analysis, 12(4), 617-629.

[69] Povh, J., Rendl, F., & Wiegele, A. (2006). A boundary point method to solve semidefinite programs. Computing, 78(3), 277-286.

[70] Qi, L., & Sun, J. (1993). A nonsmooth version of Newton's method. Mathematical programming, 58(1-3), 353-367.

[71] Qi, L. (1993). Convergence analysis of some algorithms for solving nonsmooth equations. Mathematics of operations research, 18(1), 227-244.

[72] Qi, L., & Sun, D. (1999). A survey of some nonsmooth equations and smoothing Newton methods. In Progress in optimization (pp. 121-146). Springer, Boston, MA.

[73] Rademacher, H. (1919). Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln und über die Transformation der Doppelintegrale. Mathematische Annalen, 79(4), 340-359.

[74] Web page with software of the working group of Franz Rendl: https://www.math.aau.at/or/Software/

[75] Slavakis, K., Kopsinis, Y., & Theodoridis, S. (2010, March). Adaptive algorithm for sparse system identification using projections onto weighted 1 balls. In Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on (pp. 3742-3745). IEEE.

[76] Shapiro, A. (2016). Differentiability properties of metric projections onto convex sets. Journal of Optimization Theory and Applications, 169(3), 953-964.

[77] Sturm, J. F. (2002). Implementation of interior point methods for mixed semidefinite and second order cone optimization problems. Optimization Methods and Software, 17(6), 1105-1154.

[78] Sturm, J. F. (1999). Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. Optimization methods and software, 11(1-4), 625-653.

[79] Sun, D., Toh, K. C., & Yang, L. (2014). A convergent proximal alternating direction method of multipliers for conic programming with 4-block constraints. Technical report.

[80] Sun, D., & Sun, J. (2002). Semismooth matrix-valued functions. Mathematics of Operations Research, 27(1), 150-169.

[81] Schwartz, J. T. (1969). Nonlinear functional analysis (Vol. 4). CRC Press.

[82] Sweetser, T. H. (1977). A minimal set-valued strong derivative for vector-valued Lipschitz functions. Journal of Optimization Theory and Applications, 23(4), 549-562.

[83] Taylor, A. B., Hendrickx, J. M., Glineur, F. (2017). Smooth strongly convex interpolation and exact worst-case performance of first-order methods. Mathematical Programming, 161(1-2), 307-345.

[84] Tütüncü, R. H., Toh, K. C., & Todd, M. J. (2003). Solving semidefinite-quadratic-linear programs using SDPT3. Mathematical programming, 95(2), 189-217.

[85] Todd, M. J., Toh, K. C., & Tütüncü, R. H. (1998). On the Nesterov–Todd Direction in Semidefinite Programming. SIAM Journal on Optimization, 8(3), 769-796.

[86] Toh, K.C., Todd, M.J., & Tütüncü, R.H. (1999). SDPT3 — a Matlab software package for semidefinite programming, Optimization Methods and Software, 11, 545–581.

[87] Tomov, S., Dongarra, J., & Baboulin, M. (2010). Towards dense linear algebra for hybrid GPU accelerated manycore systems. Parallel Computing, 36(5-6), 232-240.

[88] Toh, K. C. (2004). Solving large scale semidefinite programs via an iterative solver on the augmented systems. SIAM Journal on Optimization, 14(3), 670-698.

[89] Ulbrich, M. (2011). Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces (Vol. 11). SIAM.

[90] Vaisman J. (2005). Convergencia fuerte del mtodo de medias sucesi- vas para operadores lineales no-expansivos, Memoria de Ingeniera Civil Matemtica, Universidad de Chile.

[91] Wen, Z., Goldfarb, D., & Yin, W. (2010). Alternating direction augmented Lagrangian methods for semidefinite programming. Mathematical Programming Computation, 2(3-4), 203-230.

[92] Wittmann, R. (1992). Approximation of fixed points of nonexpansive mappings. Archiv der Mathematik, 58(5), 486-491.

[93] Wright, S. J. (1997). Primal-dual interior-point methods (Vol. 54). Siam.

[94] Woodbury, M. A. (1950). Inverting modified matrices. Memorandum report, 42(106), 336.

[95] Xu, H. K. (2002). Iterative algorithms for nonlinear operators. Journal of the London Mathematical Society, 66(1), 240-256.

[96] Yamashita, M., Fujisawa, K., & Kojima, M. (2003). Implementation and evaluation of SDPA 6.0 (semidefinite programming algorithm 6.0). Optimization Methods and Software, 18(4), 491-505.

[97] Yang, L., Sun, D., & Toh, K. C. (2014). SDPNAL +: A Majorized Semismooth Newton-CG Augmented Lagrangian Method for Semidefinite Programming with Nonnegative Constraints. arXiv preprint arXiv:1406.0942 .

[98] Zarantonello, E. H. (1971). Projections on Convex Sets in Hilbert Space and Spectral Theory: Part I. Projections on Convex Sets: Part II. Spectral Theory. In Contributions to nonlinear functional analysis (pp. 237-424). 1971, pp. 237-424

[99] Zhao, X. Y., Sun, D., & Toh, K. C. (2010). A Newton-CG augmented Lagrangian method for semidefinite programming. SIAM Journal on Optimization, 20(4), 1737-1765.

9 Appendix

9.1 Proof of Proposition 4.8

We still need to prove proposition 4.8. As in section 4.1 we again shift our index by one, i.e. our initial point is not $Z^{(0)}$ but $Z^{(1)}$. We define the KM iterates as usual

$$Z^{(i+1)} := Z^{(i)} - 2t_i\mathcal{H}[\nabla f(Z^{(i)})]$$

for $i \in \{1, .., k-1\}$ and see that the shifted statement

$$|(\tfrac{1}{k}\sum_{i=1}^{k} f(Z^{(i)})) - f(Z^{(*)})| \leq \tfrac{1}{k}\frac{\left\|Z^{(1)} - Z^{(*)}\right\|^2}{8t(1-t)} \tag{322}$$

is equivalent to the claim of proposition 4.8 for $t_i \equiv t \in [\frac{1}{2}, 1)$. Since this is the Appendix, it actually makes a lot of sense to include a construction for arbitrary $t_i \in (0,1)$. Let us start by defining the following quantities

$$\begin{aligned}
R &:= \left\|Z^{(1)} - Z^{(*)}\right\|, \quad & R_H &:= \langle Z^{(1)} - Z^{(*)}, \mathcal{H}[Z^{(1)} - Z^{(*)}]\rangle, \\
g_i &:= \mathcal{H}[\nabla f(Z^{(i)})] \quad \forall i \in \{1, .., k\}, & & \\
a &= ((\langle \mathcal{H}[g_i], Z^{(1)} - Z^{(*)}\rangle))_{i=1}^k \in \mathbb{R}^k, & b &= ((\langle g_i, Z^{(1)} - Z^{(*)}\rangle))_{i=1}^k \in \mathbb{R}^k \\
A &= ((\langle g_i, g_j\rangle))_{i,j=1}^k \in \mathbb{S}_+^k, & B &= ((\langle \mathcal{H}[g_i], g_j\rangle))_{i,j=1}^k \in \mathbb{S}^k \\
f_i &:= f(Z^{(k)}) \quad \forall i \in \{1, .., k\} & & \\
f_* &:= 0 & g_* &:= \nabla f(Z^{(*)}) = 0.
\end{aligned}$$

Then the easiest way to think of the upcoming analysis is to think of g_i as vectors and to think of the standard scalar product. If this was true, we could easily see the following equations

$$\begin{aligned}
G &:= [g_1, \dots, g_k] \\
A &= G^T G \\
B &= G^T \mathcal{H} G
\end{aligned} \tag{323}$$

and

$$\begin{pmatrix} (Z^{(1)} - Z^{(*)})^T \\ (\mathcal{H}[Z^{(1)} - Z^{(*)}])^T \\ G^T \\ G^T\mathcal{H} \end{pmatrix} \begin{pmatrix} (Z^{(1)} - Z^{(*)})^T \\ (\mathcal{H}[Z^{(1)} - Z^{(*)}])^T \\ G^T \\ G^T\mathcal{H} \end{pmatrix}^T = \begin{pmatrix} R^2 & R_H & b^T & a^T \\ R_H & R^2 & a^T & b^T \\ b & a & A & B \\ a & b & B & A \end{pmatrix} \in \mathbb{S}_+^{2k+2} \tag{324}$$

as well as

$$GL = (Z^{(j)} - Z^{(1)})_{j=1}^k$$

$$\mathcal{H}GL = (\mathcal{H}[Z^{(j)} - Z^{(1)})_{j=1}^k$$

for

$$L := \begin{pmatrix} 0 & -2t_1 & -2t_1 & \dots & -2t_1 \\ 0 & 0 & -2t_2 & \dots & -2t_2 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & -2t_{k-1} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{k \times k} \tag{325}$$

to be true. However in general g_i are not vectors (they may for example be matrices) and the (symmetric) scalar product is not the standard scalar product. We therefore have to work with the (symmetric and positive semidefinite) Gramian matrix

$$\begin{pmatrix} R^2 & R_H & b^T & a^T \\ R_H & R^2 & a^T & b^T \\ b & a & A & B \\ a & b & B & A \end{pmatrix} \in \mathbb{S}_+^{2k+2}$$

(or parts of it) directly. Especially the following equalities

$$AL = (\langle g_i, Z^{(j)} - Z^{(1)} \rangle)_{i,j=1}^k$$

$$L^T AL = (\langle Z^{(i)} - Z^{(1)}, Z^{(j)} - Z^{(1)} \rangle)_{i,j=1}^k$$

$$BL = (\langle \mathcal{H}[g_i], Z^{(j)} - Z^{(1)} \rangle)_{i,j=1}^k$$

$$L^T BL = (\langle \mathcal{H}[Z^{(i)} - Z^{(1)}], Z^{(j)} - Z^{(1)} \rangle)_{i,j=1}^k$$

$$ae^T + BL = (\langle \mathcal{H}[g_i], Z^{(j)} - Z^{(*)} \rangle)_{i,j=1}^k$$

$$R_H ee^T + L^T ae^T + ea^T L + L^T BL = (\langle \mathcal{H}[Z^{(i)} - Z^{(*)}], Z^{(j)} - Z^{(*)} \rangle)_{i,j=1}^k$$

$$be^T + AL = (\langle g_i, Z^{(j)} - Z^{(*)} \rangle)_{i,j=1}^k$$

$$R^2 ee^T + L^T be^T + eb^T L + L^T AL = (\langle Z^{(i)} - Z^{(*)}, Z^{(j)} - Z^{(*)} \rangle)_{i,j=1}^k$$

$$\operatorname{diag}(AL)e^T - AL = (\langle g_i, Z^{(i)} - Z^{(j)} \rangle)_{i,j=1}^k$$

$$\operatorname{diag}(BL)e^T - BL = (\langle \mathcal{H}[g_i], Z^{(i)} - Z^{(j)} \rangle)_{i,j=1}^k$$

$$\operatorname{diag}(L^T AT)e^T + e \ \operatorname{diag}(L^T AL)^T - 2L^T AL = (\langle Z^{(i)} - Z^{(j)}, Z^{(i)} - Z^{(j)} \rangle)_{i,j=1}^k = (\left\| Z^{(i)} - Z^{(j)} \right\|^2)_{i,j=1}^k$$

$$\operatorname{diag}(L^T BL)e^T + e \ \operatorname{diag}(L^T BL)^T - 2L^T BL = (\langle \mathcal{H}[Z^{(i)} - Z^{(j)}], Z^{(i)} - Z^{(j)} \rangle)_{i,j=1}^k$$

$$\operatorname{diag}(AL)e^T + e\ \operatorname{diag}(AL)^T - AL - L^TA = (\langle g_i - g_j, Z^{(i)} - Z^{(j)}\rangle)_{i,j=1}^k$$
$$\operatorname{diag}(BT)e^T + e\ \operatorname{diag}(BL)^T - BL - L^TB = (\langle \mathcal{H}[g_i - g_j], Z^{(i)} - Z^{(j)}\rangle)_{i,j=1}^k$$

$$\operatorname{diag}(A)e^T + e\ \operatorname{diag}(A)^T - 2A = (\|g_i - g_j\|^2)_{i,j=1}^k$$
$$\operatorname{diag}(B)e^T + e\ \operatorname{diag}(B)^T - 2B = (\langle \mathcal{H}[g_i - g_j], g_i - g_j\rangle)_{i,j=1}^k$$

can be verified. Note that from firm nonexpansiveness ($\|g_i - g_j\|^2 \leq \langle g_i - g_j, Z^{(i)} - Z^{(j)}\rangle$) we get the componentwise inequalities

$$\operatorname{diag}(A)e^T + e\ \operatorname{diag}(A)^T - 2A \leq \operatorname{diag}(AL)e^T + e\ \operatorname{diag}(AL)^T - AL - L^TA$$

of which only $\frac{k^2-k}{2}$ are non redundant. From $g_* = 0$ we get another k inequalities

$$\operatorname{diag}(A) \leq b + \operatorname{diag}(AL)$$

via $\|g_i\|^2 \leq \langle g_i, Z^{(i)} - Z^{(*)}\rangle$. For the actual proof we assume $t_k \equiv t \in (0,1)\ \forall k \in \mathbb{N}$. We see that in this case $\frac{L+L^T}{2} = t(I - ee^T)$ holds true. Now the main inequalities to be used here are the ones from (45),

$$\begin{aligned} &\tfrac{1}{2}\langle(\mathcal{I} - \mathcal{H})(\nabla f(X) - \nabla f(Y)), X - Y\rangle + \tfrac{1}{2}\|\nabla f(X) - \nabla f(Y)\|^2 \\ \leq\ & f(X) - f(Y) - \langle \nabla f(Y), X - Y\rangle \\ \leq\ & \tfrac{1}{2}\langle(\mathcal{I} + \mathcal{H})(\nabla f(X) - \nabla f(Y)), X - Y\rangle - \tfrac{1}{2}\|\nabla f(X) - \nabla f(Y)\|^2 \end{aligned} \tag{326}$$

that allow (for $Y = Z^{(*)}$) the following

$$\begin{aligned} &\sum_{i=1}^k (f(Z^{(i)}) - f(Z^{(*)})) \leq \tfrac{1}{2}\sum_{i=1}^k (\langle g_i, Z^{(i)} - Z^{(*)}\rangle + \langle \mathcal{H}[g_i], Z^{(i)} - Z^{(*)}\rangle - \|g_i\|^2) \\ =&\tfrac{1}{2}(e^Ta + e^T\ \operatorname{diag}(BL) + e^Tb + e^T\ \operatorname{diag}(AL) - trace(A)) \\ =&\tfrac{1}{2}(e^Ta + trace(B\tfrac{L+L^T}{2}) + e^Tb + trace(A\tfrac{L+L^T}{2}) - trace(A)) \\ =&\tfrac{1}{4}\begin{pmatrix} \frac{1}{t} - \frac{1}{t} & e^T & e^T \\ e & -I + t(I - ee^T) & t(I - ee^T) \\ e & t(I - ee^T) & -I + t(I - ee^T) \end{pmatrix} \bullet \begin{pmatrix} R^2 & a^T & b^T \\ a & A & B \\ b & B & A \end{pmatrix} \\ =&\tfrac{R^2}{4t} + (2t-1)\tfrac{trace(A)}{2} + \tfrac{1}{4}\begin{pmatrix} -\frac{1}{t} & e^T & e^T \\ e & -tI - tee^T & t(I - ee^T) \\ e & t(I - ee^T) & -tI - tee^T \end{pmatrix} \bullet \begin{pmatrix} R^2 & a^T & b^T \\ a & A & B \\ b & B & A \end{pmatrix} \\ =&\tfrac{R^2}{4t} + (2t-1)\tfrac{trace(A)}{2} - \tfrac{1}{4}\underbrace{\begin{pmatrix} \frac{1}{t} & -e^T & -e^T \\ -e & tI + tee^T & t(-I + ee^T) \\ -e & t(-I + ee^T) & tI + tee^T \end{pmatrix}}_{\succeq 0} \bullet \underbrace{\begin{pmatrix} R^2 & a^T & b^T \\ a & A & B \\ b & B & A \end{pmatrix}}_{\succeq 0} \\ \leq&\tfrac{R^2}{4t} + (2t-1)\tfrac{trace(A)}{2} \end{aligned} \tag{327}$$

inequality. Positive semidefiniteness of the matrix before the last inequality above follows for example by the following argument: For $t \in (0,1)$ we have $\begin{pmatrix} tI & -tI \\ -tI & tI \end{pmatrix} \succeq 0$ and by the Schur complement we can conclude $\begin{pmatrix} \frac{1}{t} & -e^T & -e^T \\ -e & tI + tee^T & t(-I + ee^T) \\ -e & t(-I + ee^T) & tI + tee^T \end{pmatrix} \succeq 0$. Now by using the opposite inequalities of (45) we see that

$$
\begin{aligned}
&\sum_{i=1}^{k}(f(Z^{(i)}) - f(Z^{(*)})) \geq \tfrac{1}{2}\sum_{i=1}^{k}(\langle \mathcal{H}[g_i], Z^{(i)} - Z^{(*)}\rangle - \langle g_i, Z^{(i)} - Z^{(*)}\rangle + \|g_i\|^2) \\
&= \tfrac{1}{2}(e^T a + e^T \ \operatorname{diag}(BL) - e^T b - e^T \ \operatorname{diag}(AL) + trace(A)) \\
&= \tfrac{1}{2}(e^T a + trace(B\tfrac{L+L^T}{2}) - e^T b - trace(A\tfrac{L+L^T}{2}) + trace(A)) \\
&= \tfrac{1}{4}\begin{pmatrix} \frac{1}{t} - \frac{1}{t} & e^T & -e^T \\ e & I - t(I - ee^T) & t(I - ee^T) \\ -e & t(I - ee^T) & I - t(I - ee^T) \end{pmatrix} \bullet \begin{pmatrix} R^2 & a^T & b^T \\ a & A & B \\ b & B & A \end{pmatrix} \\
&= -\tfrac{R^2}{4t} - (2t-1)\tfrac{trace(A)}{2} + \tfrac{1}{4}\underbrace{\begin{pmatrix} \frac{1}{t} & e^T & -e^T \\ e & tI + tee^T & t(I - ee^T) \\ -e & t(I - ee^T) & tI + tee^T \end{pmatrix}}_{\succeq 0} \bullet \underbrace{\begin{pmatrix} R^2 & a^T & b^T \\ a & A & B \\ b & B & A \end{pmatrix}}_{\succeq 0} \\
&\geq -\tfrac{R^2}{4t} + -(2t-1)\tfrac{trace(A)}{2}
\end{aligned}
\tag{328}
$$

can be lower bounded. Note the lower bound is the negative of the upper bound, so implying that $|(\frac{1}{k}\sum_{i=1}^{k} f(Z^{(i)})) - f(Z^{(*)})| \leq \frac{1}{k}(\frac{R^2}{4t} + (2t-1)\frac{trace(A)}{2}) = \frac{1}{k}\underbrace{(\frac{\|Z^{(1)} - Z^{(*)}\|^2}{4t} + \frac{(2t-1)\sum_{i=1}^{k}\|\nabla f(Z^{(i)})\|^2}{2})}_{\leq \frac{\|Z^{(1)} - Z^{(*)}\|^2}{8t(1-t)}}$ holds true, where we used the inequality $\sum_{i=1}^{k}\|\nabla f(Z^{(i)})\|^2 \leq \frac{\|Z^{(1)} - Z^{(*)}\|^2}{4t(1-t)}$ from the proof of proposition 4.2. This concludes the proof of proposition 4.8.

9.2 Detailed Numerical Results

Here we include details about the numerical results of the final implementation discussed in section six for sake of completeness.

	$\frac{1}{2}(n_s^2+n_s)$	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	14835	908	765	-7.74260e+00	-7.74260e+00	9.3e-08	4.2e-08	2.1e-08	1.6	6.2
brock200-2.clq	20100	9877	99	40	-1.47056e+01	-1.47056e+01	7.0e-08	3.8e-07	3.4e-07	0.3	1.1
brock200-3.clq	20100	12049	122	69	-1.12131e+01	-1.12131e+01	8.1e-08	2.5e-07	1.4e-08	0.4	1.5
brock200-4.clq	20100	13090	128	68	-9.86692e+00	-9.86692e+00	8.2e-08	1.9e-07	6.6e-08	0.3	1.2
brock400-1.clq	80200	59724	147	60	-1.03883e+01	-1.03883e+01	6.0e-08	1.3e-07	7.6e-08	1.4	4.9
brock400-2.clq	80200	59787	150	73	-1.04081e+01	-1.04081e+01	6.2e-08	1.3e-07	1.1e-07	1.4	4.7
brock400-3.clq	80200	59682	150	59	-1.04372e+01	-1.04372e+01	6.0e-08	1.3e-07	1.1e-07	1.5	5.1
brock400-4.clq	80200	59766	152	63	-1.04333e+01	-1.04333e+01	6.1e-08	1.3e-07	1.0e-07	1.5	5.1
brock800-1.clq	320400	207506	113	13	-1.92331e+01	-1.92331e+01	4.1e-08	1.3e-07	-1.0e-07	5.7	20.1
brock800-2.clq	320400	208167	113	19	-1.91281e+01	-1.91281e+01	4.2e-08	1.3e-07	-1.1e-07	5.5	19.3
brock800-3.clq	320400	207334	113	35	-1.92686e+01	-1.92686e+01	4.0e-08	1.2e-07	-9.5e-08	5.1	17.8
brock800-4.clq	320400	207644	113	38	-1.91803e+01	-1.91803e+01	3.9e-08	1.1e-07	-1.2e-07	5.0	17.4
c-fat200-1.clq	20100	1535	3566	3280	-1.84666e+01	-1.84666e+01	8.8e-08	4.4e-07	-1.2e-09	4.2	16.6
c-fat200-2.clq	20100	3236	911	772	-9.00001e+00	-8.99999e+00	2.9e-08	1.2e-06	1.4e-06	1.1	4.2
c-fat200-5.clq	20100	8474	329	276	-3.31768e+00	-3.31766e+00	4.4e-08	8.8e-07	2.3e-06	0.4	1.5
c-fat500-1.clq	125250	4460	4485	3368	-4.00000e+01	-4.00000e+01	1.5e-08	1.3e-06	-2.9e-09	35.3	120.0
c-fat500-10.clq	125250	46628	446	378	-4.00003e+00	-4.00000e+00	4.1e-08	7.6e-07	3.0e-06	3.0	9.5
c-fat500-2.clq	125250	9140	1540	1122	-2.00001e+01	-2.00000e+01	5.6e-08	4.9e-07	2.6e-06	11.5	39.1
c-fat500-5.clq	125250	23192	870	752	-7.99995e+00	-7.99999e+00	4.7e-08	7.7e-07	-2.4e-06	5.6	18.1
hamming10-2.clq	524800	518657	9834	7593	-2.00004e+00	-2.00000e+00	4.2e-08	2.2e-08	8.9e-06	436.4	1262.9
hamming10-4.clq	524800	434177	1515	1019	-2.00000e+01	-2.00000e+01	2.5e-08	4.6e-07	-4.9e-07	90.8	289.3
hamming6-2.clq	2080	1825	1171	1014	-2.00000e+00	-2.00000e+00	2.8e-08	4.1e-07	-1.5e-07	0.4	0.8
hamming6-4.clq	2080	705	1289	920	-1.20000e+01	-1.20000e+01	1.4e-07	4.4e-07	1.3e-06	0.5	2.1
hamming8-2.clq	32896	31617	3239	2702	-1.99998e+00	-2.00000e+00	8.1e-08	5.8e-08	-4.4e-06	5.8	22.9
hamming8-4.clq	32896	20865	1411	847	-1.60000e+01	-1.60000e+01	3.1e-08	7.7e-07	9.4e-08	3.3	13.2
johnson16-2-4.clq	7260	5461	1100	926	-1.50000e+01	-1.50000e+01	6.0e-08	5.7e-07	1.7e-07	0.9	3.6
johnson32-2-4.clq	123256	107881	2228	1763	-3.10000e+01	-3.10000e+01	5.8e-08	1.2e-07	8.6e-07	16.1	51.8
johnson8-2-4.clq	406	211	571	503	-6.99999e+00	-7.00001e+00	1.1e-07	7.4e-07	-8.5e-07	0.2	0.3
johnson8-4-4.clq	2485	1856	657	495	-4.99999e+00	-5.00000e+00	1.4e-07	1.1e-07	-1.1e-06	0.3	1.1
keller4.clq	14706	9436	638	337	-1.50000e+01	-1.50000e+01	9.7e-09	7.9e-07	1.5e-06	0.8	3.3
keller5.clq	301476	225991	1344	667	-3.10000e+01	-3.09999e+01	4.6e-08	2.3e-07	1.5e-06	35.7	117.7
keller6.clq	5649841	4619899	3550	963	-6.30001e+01	-6.30003e+01	1.7e-08	4.3e-07	-1.7e-06	9478.9	35698.9
p-hat1000-1.clq	500500	122254	3206	612	-9.07681e+01	-9.07681e+01	4.3e-08	1.6e-08	7.7e-09	260.7	924.7
p-hat1000-2.clq	500500	244800	64889		-6.16475e+01	-6.16475e+01	4.3e-08	8.2e-10	9.8e-08	5852.9	21104.6
p-hat1000-3.clq	500500	371747	2418		-1.82310e+01	-1.82310e+01	4.3e-08	1.2e-08	3.0e-08	211.1	751.2
p-hat1500-1.clq	1125750	284924	2382	319	-1.10996e+02	-1.10996e+02	3.5e-08	2.6e-08	3.6e-09	633.8	2312.5
p-hat1500-2.clq	1125750	568961	65543		-7.43815e+01	-7.43815e+01	3.6e-08	7.2e-10	9.8e-08	18423.4	68682.7
p-hat1500-3.clq	1125750	847245	2307		-2.15242e+01	-2.15242e+01	3.6e-08	6.7e-09	2.2e-08	628.1	2323.5
p-hat300-1.clq	45150	10934	4049	1201	-4.45819e+01	-4.45819e+01	7.7e-08	1.0e-08	1.3e-08	23.5	90.4
p-hat300-2.clq	45150	21929	39756		-2.91674e+01	-2.91674e+01	7.7e-08	1.4e-09	8.5e-08	272.0	1058.4
p-hat300-3.clq	45150	33391	2350		-1.04756e+01	-1.04756e+01	7.7e-08	1.5e-08	3.5e-08	16.1	62.9
p-hat500-1.clq	125250	31570	1808	858	-5.80355e+01	-5.80355e+01	6.0e-08	3.2e-08	1.1e-08	23.0	83.4
p-hat500-2.clq	125250	62947	57301		-3.94722e+01	-3.94722e+01	6.1e-08	8.2e-10	9.6e-08	972.1	3634.2
p-hat500-3.clq	125250	93801	2872		-1.28939e+01	-1.28939e+01	6.0e-08	3.9e-09	3.5e-08	47.9	178.9
p-hat700-1.clq	245350	61000	4960	888	-7.44577e+01	-7.44577e+01	5.1e-08	1.0e-08	1.0e-08	162.7	583.2
p-hat700-2.clq	245350	121729	63686		-5.25017e+01	-5.25016e+01	5.2e-08	7.2e-10	9.4e-08	2280.3	8326.7
p-hat700-3.clq	245350	183011	2802		-1.53169e+01	-1.53169e+01	5.1e-08	8.3e-09	3.6e-08	99.3	356.4
san1000.clq	500500	250501	7528	4854	-6.70001e+01	-6.70002e+01	3.3e-08	6.5e-07	-9.7e-07	378.6	1209.4
san200-0.7-1.clq	20100	13931	3904	2679	-9.04974e+00	-9.04973e+00	4.7e-09	7.7e-07	6.3e-07	7.4	29.5
san200-0.7-2.clq	20100	13931	1950	1232	-1.20000e+01	-1.19999e+01	8.3e-09	7.6e-07	2.7e-06	3.4	13.5
san200-0.9-1.clq	20100	17911	20751		-4.02410e+00	-4.02410e+00	9.3e-08	8.2e-09	3.7e-07	58.7	234.7
san200-0.9-2.clq	20100	17911	1004	298	-4.30710e+00	-4.30709e+00	8.9e-08	1.3e-07	5.0e-07	2.5	10.0
san200-0.9-3.clq	20100	17911	4430		-5.00000e+00	-5.00000e+00	9.3e-08	1.1e-08	9.9e-08	12.3	49.0
san400-0.5-1.clq	80200	39901	506	386	-3.42380e+01	-3.42380e+01	1.7e-08	9.4e-07	1.3e-07	3.7	12.7
san400-0.7-1.clq	80200	55861	492	167	-1.26410e+01	-1.26410e+01	6.5e-08	1.1e-07	6.1e-07	4.8	17.2
san400-0.7-2.clq	80200	55861	35660	26609	-1.50000e+01	-1.49998e+01	4.0e-08	6.2e-07	7.6e-06	207.7	698.2
san400-0.7-3.clq	80200	55861	9096	7069	-1.90000e+01	-1.89999e+01	6.1e-08	3.3e-07	2.3e-06	53.3	173.9
san400-0.9-1.clq	80200	71821	884	367	-5.20718e+00	-5.20717e+00	3.6e-08	3.7e-07	9.3e-07	7.7	27.5
sanr200-0.7.clq	20100	13869	271	84	-8.86497e+00	-8.86497e+00	9.0e-08	4.3e-08	2.6e-08	0.7	3.0
sanr200-0.9.clq	20100	17864	2273	677	-4.47872e+00	-4.47872e+00	9.3e-08	3.7e-09	4.2e-08	5.7	22.8
sanr400-0.5.clq	80200	39985	100	22	-2.03044e+01	-2.03044e+01	4.2e-08	5.5e-07	1.9e-07	1.2	4.0
sanr400-0.7.clq	80200	55870	130	49	-1.20403e+01	-1.20403e+01	6.1e-08	1.5e-07	1.3e-07	1.3	4.5

Table 12: Algorithm-Performance Table-Lovasz-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=1e-06

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	1120	908	-7.71968e+00	-7.71968e+00	9.3e-08	3.9e-08	3.7e-08	2.3	9.2
brock200-2.clq	20100	10024	19901	141	69	-1.46054e+01	-1.46054e+01	8.3e-08	2.0e-07	3.1e-08	0.5	2.1
brock200-3.clq	20100	7852	19901	167	92	-1.11618e+01	-1.11618e+01	8.6e-08	1.6e-07	5.6e-08	0.5	1.8
brock200-4.clq	20100	6811	19901	183	64	-9.82463e+00	-9.82463e+00	8.6e-08	1.3e-07	6.9e-08	0.5	2.1
brock400-1.clq	80200	20077	79801	194	45	-1.03575e+01	-1.03575e+01	6.2e-08	1.0e-07	1.1e-07	2.2	7.8
brock400-2.clq	80200	20014	79801	202	3	-1.03755e+01	-1.03755e+01	6.3e-08	9.8e-08	1.1e-07	2.6	9.3
brock400-3.clq	80200	20119	79801	200	77	-1.04060e+01	-1.04060e+01	6.2e-08	9.9e-08	1.1e-07	2.0	7.4
brock400-4.clq	80200	20035	79801	198	102	-1.03995e+01	-1.03994e+01	6.3e-08	1.0e-07	9.9e-08	1.8	6.7
brock800-1.clq	320400	112095	319601	148	44	-1.91430e+01	-1.91430e+01	4.2e-08	1.4e-07	9.1e-08	7.1	24.8
brock800-2.clq	320400	111434	319601	147	58	-1.90365e+01	-1.90365e+01	4.4e-08	1.4e-07	7.5e-08	6.7	23.0
brock800-3.clq	320400	112267	319601	149	33	-1.91785e+01	-1.91785e+01	4.2e-08	1.4e-07	7.0e-08	7.4	26.0
brock800-4.clq	320400	111957	319601	146	27	-1.90932e+01	-1.90932e+01	4.3e-08	1.4e-07	9.2e-08	7.4	26.0
c-fat200-1.clq	20100	18366	19901	4293	3957	-1.84666e+01	-1.84666e+01	8.1e-08	5.4e-07	-4.9e-09	6.5	25.7
c-fat200-2.clq	20100	16665	19901	1238	1046	-9.00003e+00	-9.00001e+00	7.7e-08	5.8e-07	1.3e-06	1.8	7.0
c-fat200-5.clq	20100	11427	19901	475	417	-3.31769e+00	-3.31767e+00	9.0e-08	9.6e-08	3.6e-06	0.7	2.6
c-fat500-1.clq	125250	120291	124751	3805	2569	-4.00001e+01	-4.00000e+01	5.6e-08	4.0e-07	1.3e-06	39.3	141.2
c-fat500-10.clq	125250	78123	124751	629	557	-4.00004e+00	-4.00000e+00	5.5e-08	3.4e-07	4.5e-06	4.8	16.9
c-fat500-2.clq	125250	115611	124751	2267	1791	-2.00000e+01	-2.00000e+01	5.5e-09	1.1e-06	-1.1e-06	20.5	72.4
c-fat500-5.clq	125250	101559	124751	1261	1088	-7.99997e+00	-8.00001e+00	2.9e-08	9.0e-07	-2.4e-06	10.2	35.8
hamming10-2.clq	524800	5120	523777	14482	11534	-1.99996e+00	-2.00000e+00	3.6e-08	6.3e-08	-7.7e-06	628.4	1787.8
hamming10-4.clq	524800	89600	523777	2201	1578	-1.99999e+01	-2.00000e+01	4.3e-08	1.3e-08	-1.7e-06	134.1	426.1
hamming6-2.clq	2080	192	2017	1817	1541	-1.99999e+00	-2.00000e+00	1.6e-07	3.3e-08	-2.3e-06	0.7	1.4
hamming6-4.clq	2080	1312	2017	1847	1455	-1.20000e+01	-1.20000e+01	4.0e-08	8.7e-07	9.0e-07	0.9	3.4
hamming8-2.clq	32896	1024	32641	4931	4093	-1.99998e+00	-2.00000e+00	7.3e-08	9.8e-08	-4.1e-06	9.4	37.1
hamming8-4.clq	32896	11776	32641	2081	1280	-1.60000e+01	-1.60000e+01	6.1e-08	4.6e-07	-5.5e-07	6.0	23.9
johnson16-2-4.clq	7260	1680	7141	1670	1431	-1.50000e+01	-1.50000e+01	8.4e-08	3.8e-07	6.6e-08	1.4	5.7
johnson32-2-4.clq	123256	14880	122761	3375	2719	-3.10000e+01	-3.10000e+01	5.3e-08	1.9e-07	-9.4e-07	25.7	92.1
johnson8-2-4.clq	406	168	379	858	773	-6.99999e+00	-7.00000e+00	2.2e-07	9.3e-08	-7.8e-07	0.3	0.5
johnson8-4-4.clq	2485	560	2416	986	752	-5.00000e+00	-4.99999e+00	4.7e-08	5.0e-07	9.0e-08	0.5	1.9
keller4.clq	14706	5100	14536	796	474	-1.50000e+01	-1.50000e+01	5.1e-08	5.6e-07	-1.8e-06	1.1	4.5
keller5.clq	301476	74710	300701	1667	876	-3.10000e+01	-3.10001e+01	1.3e-08	5.4e-07	-2.4e-06	46.7	155.8
keller6.clq	5649841	1026582	5646481	4353	1274	-6.30001e+01	-6.30002e+01	2.1e-08	2.4e-07	-7.8e-07	11448.0	42893.9
p-hat1000-1.clq	500500	377247	499501	8292	3322	-8.94781e+01	-8.94781e+01	4.3e-08	4.5e-09	1.5e-08	704.4	2437.1
p-hat1000-2.clq	500500	254701	499501	84982		-6.08624e+01	-6.08624e+01	4.3e-08	6.1e-10	1.7e-07	8292.7	29540.1
p-hat1000-3.clq	500500	127754	499501	2323		-1.81366e+01	-1.81366e+01	4.3e-08	6.4e-09	3.1e-08	207.2	734.2
p-hat1500-1.clq	1125750	839327	1124251	5880	1900	-1.09265e+02	-1.09265e+02	3.6e-08	5.9e-09	1.6e-08	1539.1	5474.9
p-hat1500-2.clq	1125750	555290	1124251	83479		-7.33087e+01	-7.33086e+01	3.6e-08	6.3e-10	1.7e-07	24939.4	92015.6
p-hat1500-3.clq	1125750	277006	1124251	2646		-2.14138e+01	-2.14138e+01	3.6e-08	7.8e-09	3.2e-08	731.7	2691.9
p-hat300-1.clq	45150	33917	44851	6252	74	-4.40679e+01	-4.40679e+01	7.7e-08	4.8e-09	2.0e-08	47.5	183.4
p-hat300-2.clq	45150	22922	44851	34505		-2.89008e+01	-2.89008e+01	7.7e-08	1.9e-09	1.3e-07	254.9	989.6
p-hat300-3.clq	45150	11460	44851	2857		-1.04261e+01	-1.04261e+01	7.7e-08	7.1e-09	3.1e-08	19.7	77.4
p-hat500-1.clq	125250	93181	124751	6373	3279	-5.72741e+01	-5.72741e+01	6.0e-08	2.2e-08	1.6e-08	91.7	341.6
p-hat500-2.clq	125250	61804	124751	72103		-3.90749e+01	-3.90749e+01	6.1e-08	6.8e-10	1.6e-07	1308.9	5030.3
p-hat500-3.clq	125250	30950	124751	2037		-1.28265e+01	-1.28265e+01	6.0e-08	2.6e-08	3.0e-08	34.5	132.1
p-hat700-1.clq	245350	183651	244651	9643	1841	-7.34971e+01	-7.34971e+01	5.1e-08	3.4e-09	2.3e-08	357.9	1282.9
p-hat700-2.clq	245350	122922	244651	96758		-5.20063e+01	-5.20063e+01	5.2e-08	5.8e-10	1.7e-07	3785.5	13794.2
p-hat700-3.clq	245350	61640	244651	3553		-1.52407e+01	-1.52407e+01	5.1e-08	8.6e-09	4.7e-08	127.3	463.2
san1000.clq	500500	249000	499501	10350	6641	-6.70001e+01	-6.69998e+01	2.7e-08	6.4e-07	2.3e-06	582.0	1866.5
san200-0.7-1.clq	20100	5970	19901	12535	10704	-9.01999e+00	-9.01997e+00	2.4e-09	6.3e-07	1.0e-06	22.9	91.2
san200-0.7-2.clq	20100	5970	19901	3010	2086	-1.20000e+01	-1.20000e+01	6.0e-08	4.8e-07	-1.5e-06	5.5	22.0
san200-0.9-1.clq	20100	1990	19901	22876		-4.01912e+00	-4.01911e+00	9.1e-08	8.2e-08	6.2e-07	67.7	270.6
san200-0.9-2.clq	20100	1990	19901	1371	778	-4.30140e+00	-4.30140e+00	9.2e-08	3.1e-08	4.2e-07	2.8	11.3
san200-0.9-3.clq	20100	1990	19901	5380		-5.00000e+00	-5.00000e+00	9.3e-08	9.2e-09	1.1e-07	15.5	61.9
san400-0.5-1.clq	80200	39900	79801	604	326	-3.42059e+01	-3.42059e+01	1.8e-08	7.7e-07	1.3e-07	5.8	21.0
san400-0.7-1.clq	80200	23940	79801	611	246	-1.26148e+01	-1.26148e+01	6.5e-08	8.0e-08	6.2e-07	5.9	21.9
san400-0.7-2.clq	80200	23940	79801	51575	40599	-1.50000e+01	-1.50001e+01	6.4e-08	1.9e-07	-3.6e-06	308.8	1128.0
san400-0.7-3.clq	80200	23940	79801	13520	10609	-1.90000e+01	-1.90001e+01	3.0e-08	5.6e-07	-3.2e-06	85.7	306.8
san400-0.9-1.clq	80200	7980	79801	1086	299	-5.20012e+00	-5.20011e+00	5.2e-08	2.2e-07	1.2e-06	10.8	38.9
sanr200-0.7.clq	20100	6032	19901	349	103	-8.82922e+00	-8.82922e+00	9.1e-08	3.1e-08	1.7e-08	1.1	4.2
sanr200-0.9.clq	20100	2037	19901	1203		-4.47656e+00	-4.47656e+00	9.3e-08	5.4e-09	3.2e-08	3.7	14.9
sanr400-0.5.clq	80200	39816	79801	135	52	-2.01700e+01	-2.01700e+01	5.9e-08	1.9e-07	6.3e-08	1.5	5.3
sanr400-0.7.clq	80200	23931	79801	176	90	-1.19950e+01	-1.19950e+01	6.1e-08	1.2e-07	1.0e-07	1.6	6.0

Table 13: Algorithm-Performance Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=1e-06

	$\frac{1}{2}(n_s^2+n_s)$	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5067	171	71	-2.74567e+01	-2.74566e+01	5.0e-08	8.6e-07	8.8e-07	0.5	1.8
brock200-2.clq	20100	10025	104	50	-1.42272e+01	-1.42272e+01	7.9e-08	2.8e-07	4.3e-08	0.3	1.1
brock200-3.clq	20100	7853	114	37	-1.88205e+01	-1.88205e+01	5.3e-08	6.1e-07	-7.0e-07	0.3	1.2
brock200-4.clq	20100	6812	134	58	-2.12934e+01	-2.12935e+01	6.4e-08	5.4e-07	-4.5e-07	0.3	1.3
brock400-1.clq	80200	20078	206	78	-3.97019e+01	-3.97019e+01	2.6e-08	1.0e-06	4.7e-07	2.0	6.9
brock400-2.clq	80200	20015	211	49	-3.95605e+01	-3.95606e+01	2.5e-08	1.0e-06	-4.8e-07	2.3	8.0
brock400-3.clq	80200	20120	209	78	-3.94806e+01	-3.94806e+01	5.2e-08	5.8e-07	4.7e-07	2.1	7.1
brock400-4.clq	80200	20036	200	79	-3.95996e+01	-3.95997e+01	2.7e-08	9.4e-07	-3.4e-07	1.9	6.6
brock800-1.clq	320400	112096	162	38	-4.22219e+01	-4.22219e+01	2.5e-08	7.5e-07	-8.6e-07	7.6	26.8
brock800-2.clq	320400	111435	162	51	-4.24659e+01	-4.24660e+01	1.9e-08	8.7e-07	-7.0e-07	7.3	25.5
brock800-3.clq	320400	112268	158		-4.22424e+01	-4.22423e+01	2.7e-08	7.4e-07	8.9e-07	8.2	29.6
brock800-4.clq	320400	111958	163	49	-4.23490e+01	-4.23491e+01	3.2e-08	6.1e-07	-9.3e-07	7.3	25.8
c-fat200-1.clq	20100	18367	1497	999	-1.20000e+01	-1.20000e+01	2.5e-08	3.7e-07	-1.1e-06	2.1	8.3
c-fat200-2.clq	20100	16666	18591	15981	-2.40000e+01	-2.40001e+01	8.4e-08	2.5e-07	-4.0e-07	22.3	89.1
c-fat200-5.clq	20100	11428	5709	4830	-6.03454e+01	-6.03453e+01	9.2e-08	1.5e-07	8.1e-07	6.6	26.4
c-fat500-1.clq	125250	120292	2784	1813	-1.40000e+01	-1.40000e+01	6.8e-09	2.6e-07	1.3e-06	23.2	77.0
c-fat500-10.clq	125250	78124	5369	3231	-1.26000e+02	-1.26000e+02	2.3e-08	7.9e-07	-7.5e-07	43.0	146.0
c-fat500-2.clq	125250	115612	2722	1632	-2.60000e+01	-2.59999e+01	1.3e-08	3.7e-07	8.2e-07	23.2	78.1
c-fat500-5.clq	125250	101560	4868	3120	-6.39999e+01	-6.39999e+01	5.0e-08	3.4e-07	2.6e-07	38.6	128.6
hamming10-2.clq	524800	5121	58578	52367	-5.12000e+02	-5.12001e+02	9.1e-09	1.4e-06	-1.1e-06	2137.8	6191.1
hamming10-4.clq	524800	89601	1686	1330	-5.12001e+01	-5.12000e+01	4.2e-08	1.5e-07	1.3e-06	90.0	290.9
hamming6-2.clq	2080	193	3800	3475	-3.20000e+01	-3.20001e+01	8.3e-08	1.1e-06	-1.1e-06	1.3	2.2
hamming6-4.clq	2080	1313	353	314	-5.33332e+00	-5.33333e+00	1.4e-07	2.4e-07	-5.7e-07	0.1	0.5
hamming8-2.clq	32896	1025	15565	13488	-1.28000e+02	-1.28000e+02	3.6e-08	1.2e-06	5.7e-07	25.4	101.1
hamming8-4.clq	32896	11777	726	591	-1.60000e+01	-1.60000e+01	8.0e-08	1.4e-07	-1.2e-06	1.5	6.0
johnson16-2-4.clq	7260	1681	558	446	-8.00002e+00	-7.99999e+00	6.0e-08	1.0e-06	1.4e-06	0.4	1.5
johnson32-2-4.clq	123256	14881	1041	794	-1.59999e+01	-1.60000e+01	5.9e-08	5.1e-08	-2.7e-06	7.3	24.4
johnson8-2-4.clq	406	169	302	267	-3.99999e+00	-4.00000e+00	1.1e-07	8.1e-07	-2.1e-07	0.1	0.2
johnson8-4-4.clq	2485	561	917	738	-1.40000e+01	-1.40000e+01	6.5e-08	1.1e-06	2.9e-07	0.4	1.5
keller4.clq	14706	5101	172	116	-1.40123e+01	-1.40122e+01	6.6e-08	6.5e-07	9.0e-07	0.3	1.0
keller5.clq	301476	74711	227	60	-3.09999e+01	-3.10000e+01	2.8e-08	8.7e-07	-1.4e-06	7.5	26.5
keller6.clq	5649841	1026583	324	138	-6.30000e+01	-6.29999e+01	4.1e-09	1.1e-06	8.2e-07	757.9	2832.6
p-hat1000-1.clq	500500	377248	2043		-1.76081e+01	-1.76081e+01	4.3e-08	1.3e-08	2.4e-08	179.4	639.0
p-hat1000-2.clq	500500	254702	51509		-5.56071e+01	-5.56071e+01	4.3e-08	9.4e-10	9.1e-08	4655.6	16825.3
p-hat1000-3.clq	500500	127755	2746	1169	-8.48020e+01	-8.48020e+01	4.3e-08	1.1e-08	3.5e-09	200.4	695.1
p-hat1500-1.clq	1125750	839328	2296		-2.20061e+01	-2.20061e+01	3.6e-08	1.2e-08	2.7e-08	633.2	2333.0
p-hat1500-2.clq	1125750	555291	71107		-7.75573e+01	-7.75573e+01	3.6e-08	6.0e-10	9.6e-08	19729.3	73328.0
p-hat1500-3.clq	1125750	277007	3329	803	-1.15434e+02	-1.15434e+02	3.6e-08	1.8e-08	8.3e-09	816.5	2960.7
p-hat300-1.clq	45150	33918	3636		-1.00680e+01	-1.00680e+01	7.7e-08	3.1e-09	4.1e-08	24.7	96.3
p-hat300-2.clq	45150	22923	31321		-2.69660e+01	-2.69660e+01	7.7e-08	1.7e-09	7.8e-08	215.6	838.4
p-hat300-3.clq	45150	11461	3676	2624	-4.11699e+01	-4.11699e+01	7.7e-08	2.8e-08	2.3e-08	14.6	56.0
p-hat500-1.clq	125250	93182	1623		-1.30741e+01	-1.30741e+01	6.0e-08	9.9e-09	3.0e-08	27.2	101.5
p-hat500-2.clq	125250	61805	41922		-3.89747e+01	-3.89747e+01	6.1e-08	1.1e-09	8.1e-08	710.6	2674.2
p-hat500-3.clq	125250	30951	2179	1294	-5.85679e+01	-5.85679e+01	6.0e-08	2.5e-08	1.2e-08	25.4	90.6
p-hat700-1.clq	245350	183652	2310		-1.51199e+01	-1.51199e+01	5.1e-08	9.8e-09	3.9e-08	81.5	292.4
p-hat700-2.clq	245350	122923	45757		-4.90193e+01	-4.90193e+01	5.2e-08	1.2e-09	8.2e-08	1652.7	6023.4
p-hat700-3.clq	245350	61641	5402	2595	-7.27350e+01	-7.27350e+01	5.1e-08	5.2e-09	8.2e-09	150.9	522.1
san1000.clq	500500	249001	3955	1710	-1.49999e+01	-1.50000e+01	4.0e-08	4.0e-07	-4.3e-06	242.8	805.9
san200-0.7-1.clq	20100	5971	21959	18901	-3.00000e+01	-3.00002e+01	7.2e-09	1.2e-06	-3.9e-06	29.4	117.4
san200-0.7-2.clq	20100	5971	129289	7303	-1.80000e+01	-1.80000e+01	1.1e-08	1.2e-06	4.1e-07	375.7	1498.4
san200-0.9-1.clq	20100	1991	22898	19953	-7.00001e+01	-7.00001e+01	8.7e-08	4.9e-07	-4.8e-08	30.4	121.4
san200-0.9-2.clq	20100	1991	22100	19119	-6.00001e+01	-5.99999e+01	8.8e-08	4.4e-07	1.4e-06	29.8	119.0
san200-0.9-3.clq	20100	1991	301413	123803	-4.40001e+01	-4.39999e+01	7.8e-08	7.3e-07	2.2e-06	649.2	2593.1
san400-0.5-1.clq	80200	39901	35743	26698	-1.30000e+01	-1.30001e+01	6.7e-08	1.5e-07	-1.6e-06	209.1	699.5
san400-0.7-1.clq	80200	23941	41409	35104	-3.99999e+01	-3.99999e+01	6.5e-08	2.8e-07	7.3e-07	203.8	669.6
san400-0.7-2.clq	80200	23941	35972	27672	-3.00000e+01	-2.99998e+01	6.1e-08	5.2e-07	4.0e-06	197.7	662.7
san400-0.7-3.clq	80200	23941	3305	1825	-2.20000e+01	-2.20000e+01	2.0e-08	1.1e-06	1.8e-07	23.8	84.6
san400-0.9-1.clq	80200	7981	44989	38974	-1.00000e+02	-9.99995e+01	9.5e-09	1.3e-06	2.5e-06	205.9	678.6
sanr200-0.7.clq	20100	6033	147	54	-2.38362e+01	-2.38361e+01	4.8e-08	8.2e-07	7.6e-07	0.4	1.5
sanr200-0.9.clq	20100	2038	298	121	-4.92735e+01	-4.92736e+01	3.7e-08	1.1e-06	-2.2e-07	0.7	2.8
sanr400-0.5.clq	80200	39817	100	23	-2.03177e+01	-2.03177e+01	3.2e-08	6.3e-07	-3.1e-07	1.2	4.0
sanr400-0.7.clq	80200	23932	168	62	-3.42783e+01	-3.42783e+01	2.6e-08	9.1e-07	4.6e-07	1.7	5.8

Table 14: Algorithm-Performance Table-LovaszComplement-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=1e-06

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	14834	19901	179	86	-2.71967e+01	-2.71967e+01	6.6e-08	5.3e-07	4.8e-08	0.5	2.0
brock200-2.clq	20100	9876	19901	174	105	-1.41310e+01	-1.41310e+01	8.6e-08	2.0e-07	9.9e-09	0.5	1.8
brock200-3.clq	20100	12048	19901	156	86	-1.86718e+01	-1.86718e+01	7.4e-08	3.1e-07	-4.8e-08	0.4	1.6
brock200-4.clq	20100	13089	19901	159	86	-2.11211e+01	-2.11211e+01	6.5e-08	5.1e-07	-4.8e-08	0.4	1.6
brock400-1.clq	80200	59723	79801	198	82	-3.93310e+01	-3.93309e+01	4.2e-08	6.0e-07	1.1e-07	2.1	7.4
brock400-2.clq	80200	59786	79801	193	62	-3.91965e+01	-3.91964e+01	3.6e-08	7.1e-07	6.0e-07	2.2	7.8
brock400-3.clq	80200	59681	79801	198	78	-3.91604e+01	-3.91604e+01	3.4e-08	7.2e-07	-6.4e-07	2.1	7.5
brock400-4.clq	80200	59765	79801	194	32	-3.92314e+01	-3.92314e+01	4.6e-08	5.5e-07	7.3e-07	2.4	8.6
brock800-1.clq	320400	207505	319601	166	37	-4.18674e+01	-4.18673e+01	3.6e-08	4.8e-07	5.4e-07	8.7	30.3
brock800-2.clq	320400	208166	319601	166	51	-4.21043e+01	-4.21042e+01	3.4e-08	4.8e-07	5.7e-07	8.3	29.0
brock800-3.clq	320400	207333	319601	164	38	-4.18825e+01	-4.18825e+01	2.8e-08	5.9e-07	3.9e-07	8.5	29.9
brock800-4.clq	320400	207643	319601	166	31	-4.20006e+01	-4.20006e+01	3.6e-08	5.0e-07	6.1e-07	8.8	30.9
c-fat200-1.clq	20100	1534	19901	2261	1596	-1.20000e+01	-1.20000e+01	8.2e-08	1.5e-07	-2.3e-07	3.4	13.3
c-fat200-2.clq	20100	3235	19901	28263	24540	-2.40000e+01	-2.40000e+01	9.3e-08	5.1e-09	7.2e-07	36.4	145.5
c-fat200-5.clq	20100	8473	19901	8329	7236	-6.03453e+01	-6.03454e+01	3.8e-08	6.8e-07	-3.7e-07	11.9	47.3
c-fat500-1.clq	125250	4459	124751	4072	2796	-1.40000e+01	-1.40000e+01	3.0e-08	1.9e-07	-6.9e-07	35.0	117.5
c-fat500-10.clq	125250	46627	124751	6794	4283	-1.26000e+02	-1.26000e+02	1.7e-08	6.7e-07	-8.7e-07	58.8	214.5
c-fat500-2.clq	125250	9139	124751	4120	2782	-2.60000e+01	-2.60000e+01	5.8e-08	8.7e-08	9.9e-07	35.1	121.9
c-fat500-5.clq	125250	23191	124751	6216	4243	-6.40000e+01	-6.40002e+01	2.6e-08	4.5e-07	-1.1e-06	51.3	185.5
hamming10-2.clq	524800	518656	523777	88494	76427	-5.12001e+02	-5.12000e+02	4.3e-08	1.2e-07	8.1e-07	4316.0	12893.4
hamming10-4.clq	524800	434176	523777	2639	2123	-4.26668e+01	-4.26666e+01	3.9e-08	4.1e-07	1.6e-06	173.4	553.3
hamming6-2.clq	2080	1824	2017	6005	5312	-3.20000e+01	-3.20000e+01	1.5e-07	1.5e-07	8.1e-07	2.5	4.5
hamming6-4.clq	2080	704	2017	937	827	-3.99999e+00	-4.00000e+00	8.2e-08	5.5e-07	-1.0e-06	0.4	1.7
hamming8-2.clq	32896	31616	32641	22513	19670	-1.28000e+02	-1.28000e+02	5.7e-08	8.3e-07	1.2e-06	50.0	199.6
hamming8-4.clq	32896	20864	32641	1029	814	-1.60000e+01	-1.60000e+01	8.1e-08	1.2e-07	-7.9e-07	2.7	10.7
johnson16-2-4.clq	7260	5460	7141	832	681	-7.99999e+00	-7.99999e+00	4.9e-08	8.8e-07	-2.3e-07	0.7	2.7
johnson32-2-4.clq	123256	107880	122761	1537	1204	-1.59999e+01	-1.60000e+01	4.0e-08	7.9e-07	-2.3e-06	13.2	47.4
johnson8-2-4.clq	406	210	379	377	359	-4.00001e+00	-4.00000e+00	2.2e-07	8.3e-08	1.0e-06	0.1	0.2
johnson8-4-4.clq	2485	1855	2416	1248	1070	-1.40000e+01	-1.40000e+01	1.2e-07	5.9e-07	-1.8e-07	0.6	2.4
keller4.clq	14706	9435	14536	620	482	-1.34659e+01	-1.34659e+01	9.8e-08	1.1e-07	1.0e-07	0.9	3.6
keller5.clq	301476	225990	300701	5910		-3.09957e+01	-3.09957e+01	4.6e-08	3.3e-07	2.4e-08	244.5	888.7
keller6.clq	5649841	4619898	5646481	401	148	-6.30000e+01	-6.29999e+01	3.9e-09	9.7e-07	4.9e-07	1048.0	3887.8
p-hat1000-1.clq	500500	122253	499501	1897		-1.75222e+01	-1.75222e+01	4.3e-08	1.6e-08	4.2e-08	167.0	591.6
p-hat1000-2.clq	500500	244799	499501	57684		-5.48435e+01	-5.48435e+01	4.3e-08	4.9e-09	1.6e-07	5608.4	20023.1
p-hat1000-3.clq	500500	371746	499501	5553	3146	-8.35284e+01	-8.35284e+01	4.3e-08	3.8e-09	1.1e-08	435.1	1475.9
p-hat1500-1.clq	1125750	284923	1124251	2504		-2.18924e+01	-2.18924e+01	3.6e-08	9.4e-09	3.2e-08	698.6	2571.8
p-hat1500-2.clq	1125750	568960	1124251	90934		-7.64554e+01	-7.64554e+01	3.6e-08	2.7e-09	1.8e-07	27149.8	100289.9
p-hat1500-3.clq	1125750	847244	1124251	7637	931	-1.13655e+02	-1.13655e+02	3.6e-08	4.9e-09	2.0e-08	2216.7	8073.2
p-hat300-1.clq	45150	10933	44851	4506		-1.00202e+01	-1.00202e+01	7.7e-08	4.7e-09	4.0e-08	31.5	124.0
p-hat300-2.clq	45150	21928	44851	38848		-2.67137e+01	-2.67137e+01	7.7e-08	1.3e-09	1.2e-07	286.5	1115.2
p-hat300-3.clq	45150	33390	44851	4820	3434	-4.07003e+01	-4.07003e+01	7.7e-08	1.9e-08	3.1e-08	22.4	87.6
p-hat500-1.clq	125250	31569	124751	2958		-1.30079e+01	-1.30079e+01	6.0e-08	7.4e-09	3.8e-08	50.8	194.3
p-hat500-2.clq	125250	62946	124751	51263		-3.85594e+01	-3.85594e+01	6.1e-08	1.0e-09	1.4e-07	933.8	3587.1
p-hat500-3.clq	125250	93800	124751	5415	3178	-5.78110e+01	-5.78110e+01	6.1e-08	4.7e-09	1.6e-08	73.8	273.2
p-hat700-1.clq	245350	60999	244651	2856		-1.50451e+01	-1.50451e+01	5.1e-08	5.2e-09	3.1e-08	102.1	371.3
p-hat700-2.clq	245350	121728	244651	61179		-4.84400e+01	-4.84400e+01	5.2e-08	8.4e-10	1.5e-07	2392.1	8730.1
p-hat700-3.clq	245350	183010	244651	6887	3576	-7.17551e+01	-7.17551e+01	5.1e-08	4.2e-09	1.4e-08	217.8	754.8
san1000.clq	500500	250500	499501	4819	2177	-1.49998e+01	-1.50000e+01	4.3e-08	1.2e-07	-4.9e-06	321.2	1061.4
san200-0.7-1.clq	20100	13930	19901	32675	28874	-3.00000e+01	-2.99999e+01	8.1e-08	4.7e-07	2.5e-06	49.6	198.2
san200-0.7-2.clq	20100	13930	19901	12104	6859	-1.80000e+01	-1.80002e+01	5.6e-09	9.6e-07	-5.9e-06	27.2	108.5
san200-0.9-1.clq	20100	17910	19901	34299	30576	-7.00001e+01	-7.00002e+01	6.4e-08	7.9e-07	-9.0e-07	51.7	206.7
san200-0.9-2.clq	20100	17910	19901	33153	29369	-6.00000e+01	-6.00002e+01	5.2e-08	9.0e-07	-2.2e-06	51.4	205.1
san200-0.9-3.clq	20100	17910	19901	26556	17709	-4.39999e+01	-4.40001e+01	8.3e-08	5.0e-07	-2.0e-06	51.9	207.4
san400-0.5-1.clq	80200	39900	79801	52886	43255	-1.30000e+01	-1.29998e+01	5.4e-08	4.9e-07	9.2e-06	315.3	1139.9
san400-0.7-1.clq	80200	55860	79801	61721	53908	-3.99999e+01	-4.00002e+01	6.2e-08	3.8e-07	-2.9e-06	357.1	1257.5
san400-0.7-2.clq	80200	55860	79801	54095	44713	-3.00000e+01	-3.00005e+01	9.2e-09	9.5e-07	-7.5e-06	334.3	1180.1
san400-0.7-3.clq	80200	55860	79801	4725	2806	-2.20001e+01	-2.20000e+01	6.4e-08	3.0e-07	3.0e-06	37.3	135.7
san400-0.9-1.clq	80200	71820	79801	67364	59817	-1.00000e+02	-1.00000e+02	3.9e-08	8.9e-07	-1.8e-06	381.3	1350.5
sanr200-0.7.clq	20100	13868	19901	172	34	-2.36333e+01	-2.36333e+01	6.3e-08	6.0e-07	-2.8e-08	0.6	2.2
sanr200-0.9.clq	20100	17863	19901	322	149	-4.89046e+01	-4.89045e+01	4.7e-08	8.6e-07	1.3e-07	0.8	3.3
sanr400-0.5.clq	80200	39984	79801	131	55	-2.01782e+01	-2.01782e+01	6.0e-08	2.0e-07	1.4e-08	1.4	5.1
sanr400-0.7.clq	80200	55869	79801	173	62	-3.39666e+01	-3.39666e+01	4.2e-08	5.9e-07	-2.3e-07	1.9	6.8

Table 15: Algorithm-Performance Table-LovaszSchrijverComplement-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=1e-06

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	16		-8.32421e+00	-7.43329e+00	6.9e-03	2.6e-02	5.3e-02	0.1	0.3
brock200-2.clq	20100	10024	19901	19		-1.46983e+01	-1.34880e+01	5.4e-03	5.4e-02	4.1e-02	0.1	0.3
brock200-3.clq	20100	7852	19901	18		-1.21018e+01	-1.07068e+01	6.0e-03	3.4e-02	5.9e-02	0.1	0.3
brock200-4.clq	20100	6811	19901	17		-1.04237e+01	-9.34865e+00	6.6e-03	3.4e-02	5.2e-02	0.1	0.3
brock400-1.clq	80200	20077	79801	17		-1.12225e+01	-1.03706e+01	5.2e-03	2.1e-02	3.8e-02	0.2	0.8
brock400-2.clq	80200	20014	79801	17		-1.12485e+01	-1.03663e+01	5.2e-03	2.1e-02	3.9e-02	0.2	0.9
brock400-3.clq	80200	20119	79801	17		-1.13097e+01	-1.03868e+01	5.2e-03	2.1e-02	4.1e-02	0.2	0.9
brock400-4.clq	80200	20035	79801	17		-1.12995e+01	-1.03856e+01	5.2e-03	2.2e-02	4.0e-02	0.2	0.9
brock800-1.clq	320400	112095	319601	19		-2.08527e+01	-1.96542e+01	3.4e-03	3.0e-02	2.9e-02	1.1	3.9
brock800-2.clq	320400	111434	319601	19		-2.06998e+01	-1.95431e+01	3.4e-03	2.9e-02	2.8e-02	1.1	3.9
brock800-3.clq	320400	112267	319601	19		-2.08947e+01	-1.96619e+01	3.4e-03	2.9e-02	3.0e-02	1.1	3.9
brock800-4.clq	320400	111957	319601	19		-2.07853e+01	-1.96023e+01	3.4e-03	3.0e-02	2.9e-02	1.1	3.9
c-fat200-1.clq	20100	18366	19901	38		-2.37676e+01	-1.90540e+01	8.0e-03	4.4e-02	1.1e-01	0.1	0.4
c-fat200-2.clq	20100	16665	19901	24		-8.73372e+00	-8.19391e+00	5.9e-03	6.7e-02	3.0e-02	0.1	0.2
c-fat200-5.clq	20100	11427	19901	21		-4.49968e+00	-3.34391e+00	8.2e-03	8.1e-03	1.3e-01	0.1	0.2
c-fat500-1.clq	125250	120291	124751	64		-2.62973e+01	-4.26320e+01	5.1e-03	5.7e-02	-2.3e-01	0.9	3.4
c-fat500-10.clq	125250	78123	124751	21		-4.04194e+00	-4.09274e+00	4.9e-03	2.2e-02	-5.6e-03	0.3	1.1
c-fat500-2.clq	125250	115611	124751	33		-1.11332e+01	-2.04595e+01	4.9e-03	4.9e-02	-2.9e-01	0.5	1.7
c-fat500-5.clq	125250	101559	124751	21		-9.47383e+00	-8.03843e+00	5.2e-03	2.9e-02	7.8e-02	0.3	1.1
hamming10-2.clq	524800	5120	523777	22		-2.89037e+00	-2.26740e+00	2.9e-03	7.6e-03	1.0e-01	1.9	6.8
hamming10-4.clq	524800	89600	523777	26		-1.76010e+01	-2.03199e+01	3.4e-03	2.0e-02	-7.0e-02	2.4	8.7
hamming6-2.clq	2080	192	2017	23		-1.93709e+00	-2.22150e+00	1.2e-02	1.9e-02	-5.5e-02	0.0	0.1
hamming6-4.clq	2080	1312	2017	134		-9.18996e+00	-1.17619e+01	1.5e-02	1.4e-02	-1.2e-01	0.1	0.3
hamming8-2.clq	32896	1024	32641	20		-2.25279e+00	-2.17652e+00	7.0e-03	7.3e-03	1.4e-02	0.1	0.3
hamming8-4.clq	32896	11776	32641	122		-1.90314e+01	-1.43786e+01	5.5e-03	5.1e-02	1.4e-01	0.4	1.6
johnson16-2-4.clq	7260	1680	7141	213	25	-1.31484e+01	-1.60591e+01	9.9e-03	3.0e-02	-9.6e-02	0.2	0.8
johnson32-2-4.clq	123256	14880	122761	472	5	-3.50578e+01	-2.92557e+01	5.5e-03	1.6e-02	8.9e-02	5.5	20.9
johnson8-2-4.clq	406	168	379	106	29	-7.52399e+00	-6.28139e+00	1.1e-02	6.2e-02	8.4e-02	0.0	0.2
johnson8-4-4.clq	2485	560	2416	88		-5.20716e+00	-4.44908e+00	2.3e-03	5.3e-02	7.1e-02	0.1	0.2
keller4.clq	14706	5100	14536	21		-1.31688e+01	-1.42288e+01	6.0e-03	2.8e-02	-3.7e-02	0.0	0.2
keller5.clq	301476	74710	300701	32		-3.31987e+01	-3.06092e+01	3.8e-03	3.1e-02	4.0e-02	1.3	4.6
keller6.clq	5649841	1026582	5646481	45		-5.27500e+01	-5.89292e+01	2.1e-03	1.7e-02	-5.5e-02	163.8	631.3
p-hat1000-1.clq	500500	377247	499501	27		-8.84995e+01	-8.58606e+01	2.2e-03	7.4e-02	1.5e-02	2.7	9.7
p-hat1000-2.clq	500500	254701	499501	20		-5.92983e+01	-5.14508e+01	2.2e-03	6.3e-02	7.0e-02	2.1	7.2
p-hat1000-3.clq	500500	127754	499501	15		-1.69298e+01	-1.84168e+01	3.2e-03	2.5e-02	-4.1e-02	1.5	5.4
p-hat1500-1.clq	1125750	839327	1124251	27		-1.07007e+02	-1.03162e+02	1.7e-03	7.3e-02	1.8e-02	8.6	31.4
p-hat1500-2.clq	1125750	555290	1124251	20		-7.09610e+01	-6.15436e+01	1.7e-03	6.2e-02	7.1e-02	6.2	22.8
p-hat1500-3.clq	1125750	277006	1124251	15		-2.00978e+01	-2.18155e+01	2.6e-03	2.4e-02	-4.0e-02	4.7	17.2
p-hat300-1.clq	45150	33917	44851	26		-4.33551e+01	-4.47333e+01	3.8e-03	6.8e-02	-1.5e-02	0.2	0.8
p-hat300-2.clq	45150	22922	44851	18		-2.99441e+01	-2.63918e+01	4.5e-03	5.8e-02	6.2e-02	0.1	0.6
p-hat300-3.clq	45150	11460	44851	15		-9.96007e+00	-1.04555e+01	5.4e-03	2.5e-02	-2.3e-02	0.1	0.5
p-hat500-1.clq	125250	93181	124751	26		-5.62556e+01	-5.69736e+01	2.9e-03	6.9e-02	-6.3e-03	0.5	1.9
p-hat500-2.clq	125250	61804	124751	19		-3.89517e+01	-3.43815e+01	3.2e-03	5.9e-02	6.1e-02	0.4	1.4
p-hat500-3.clq	125250	30950	124751	15		-1.21191e+01	-1.29747e+01	4.3e-03	2.4e-02	-3.3e-02	0.3	1.1
p-hat700-1.clq	245350	183651	244651	27		-7.12611e+01	-6.99765e+01	2.4e-03	7.1e-02	9.0e-03	1.1	4.0
p-hat700-2.clq	245350	122922	244651	20		-4.93465e+01	-4.26048e+01	2.5e-03	6.5e-02	7.3e-02	0.8	3.0
p-hat700-3.clq	245350	61640	244651	15		-1.43403e+01	-1.54622e+01	3.6e-03	2.4e-02	-3.6e-02	0.6	2.2
san1000.clq	500500	249000	499501	30		-6.31366e+01	-5.71781e+01	2.7e-03	5.2e-02	4.9e-02	3.0	10.7
san200-0.7-1.clq	20100	5970	19901	19		-9.95035e+00	-8.08326e+00	6.0e-03	3.9e-02	9.8e-02	0.1	0.3
san200-0.7-2.clq	20100	5970	19901	19		-1.21176e+01	-1.20027e+01	6.2e-03	3.2e-02	4.6e-03	0.1	0.3
san200-0.9-1.clq	20100	1990	19901	14		-4.21037e+00	-3.96735e+00	7.3e-03	1.5e-02	2.6e-02	0.1	0.2
san200-0.9-2.clq	20100	1990	19901	16		-4.56345e+00	-4.39166e+00	7.0e-03	1.0e-02	1.7e-02	0.1	0.2
san200-0.9-3.clq	20100	1990	19901	15		-5.24576e+00	-4.67924e+00	7.5e-03	1.3e-02	5.2e-02	0.1	0.2
san400-0.5-1.clq	80200	39900	79801	30		-3.95468e+01	-3.30856e+01	5.3e-03	3.9e-02	8.8e-02	0.4	1.5
san400-0.7-1.clq	80200	23940	79801	25		-1.45764e+01	-1.30042e+01	6.0e-03	2.0e-02	5.5e-02	0.3	1.2
san400-0.7-2.clq	80200	23940	79801	19		-1.59709e+01	-1.47347e+01	4.7e-03	3.0e-02	3.9e-02	0.3	0.9
san400-0.7-3.clq	80200	23940	79801	16		-1.77139e+01	-1.60973e+01	4.2e-03	3.3e-02	4.6e-02	0.2	0.8
san400-0.9-1.clq	80200	7980	79801	16		-5.64446e+00	-5.64107e+00	4.9e-03	1.8e-02	2.8e-04	0.2	0.8
sanr200-0.7.clq	20100	6032	19901	17		-9.41591e+00	-8.57730e+00	6.5e-03	2.6e-02	4.4e-02	0.1	0.3
sanr200-0.9.clq	20100	2037	19901	16		-4.95625e+00	-4.53049e+00	7.3e-03	1.0e-02	4.1e-02	0.1	0.2
sanr400-0.5.clq	80200	39816	79801	20		-2.16337e+01	-1.88655e+01	4.1e-03	4.8e-02	6.7e-02	0.3	1.0
sanr400-0.7.clq	80200	23931	79801	18		-1.33965e+01	-1.21954e+01	4.8e-03	2.4e-02	4.5e-02	0.3	0.9

Table 16: Algorithm-Performance with increased tolerance: Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=0.1

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	36		-7.73551e+00	-7.73041e+00	7.4e-04	1.3e-03	3.1e-04	0.2	0.7
brock200-2.clq	20100	10024	19901	39		-1.46626e+01	-1.45368e+01	5.7e-04	4.3e-03	4.2e-03	0.2	0.7
brock200-3.clq	20100	7852	19901	36		-1.12415e+01	-1.11511e+01	7.2e-04	2.5e-03	3.9e-03	0.2	0.6
brock200-4.clq	20100	6811	19901	36		-9.87290e+00	-9.84497e+00	7.3e-04	2.3e-03	1.3e-03	0.1	0.6
brock400-1.clq	80200	20077	79801	37		-1.03183e+01	-1.03504e+01	5.2e-04	1.3e-03	-1.5e-03	0.5	1.8
brock400-2.clq	80200	20014	79801	37		-1.03330e+01	-1.03706e+01	5.2e-04	1.3e-03	-1.7e-03	0.5	1.8
brock400-3.clq	80200	20119	79801	37		-1.03581e+01	-1.04002e+01	5.2e-04	1.4e-03	-1.9e-03	0.5	1.8
brock400-4.clq	80200	20035	79801	37		-1.03564e+01	-1.03939e+01	5.2e-04	1.4e-03	-1.7e-03	0.5	1.8
brock800-1.clq	320400	112095	319601	38		-1.91097e+01	-1.92079e+01	3.5e-04	2.8e-03	-2.5e-03	2.1	7.6
brock800-2.clq	320400	111434	319601	38		-1.90093e+01	-1.90968e+01	3.4e-04	2.6e-03	-2.2e-03	2.2	7.6
brock800-3.clq	320400	112267	319601	38		-1.91446e+01	-1.92412e+01	3.4e-04	2.7e-03	-2.5e-03	2.1	7.6
brock800-4.clq	320400	111957	319601	38		-1.90581e+01	-1.91538e+01	3.5e-04	2.7e-03	-2.4e-03	2.1	7.6
c-fat200-1.clq	20100	18366	19901	203	1	-1.78458e+01	-1.83361e+01	8.1e-04	5.1e-03	-1.3e-02	0.5	2.1
c-fat200-2.clq	20100	16665	19901	101	7	-9.26163e+00	-9.11861e+00	5.8e-04	8.1e-03	7.4e-03	0.2	0.9
c-fat200-5.clq	20100	11427	19901	83	29	-3.26763e+00	-3.28454e+00	1.7e-04	8.2e-03	-2.2e-03	0.2	0.7
c-fat500-1.clq	125250	120291	124751	347		-4.06163e+01	-4.04576e+01	2.6e-04	1.0e-02	1.9e-03	5.3	20.0
c-fat500-10.clq	125250	78123	124751	59	5	-4.36080e+00	-4.02284e+00	5.2e-04	3.9e-03	3.6e-02	0.7	2.8
c-fat500-2.clq	125250	115611	124751	147		-2.02350e+01	-1.97641e+01	3.3e-04	8.6e-03	1.1e-02	2.0	7.4
c-fat500-5.clq	125250	101559	124751	116	15	-7.67593e+00	-7.88754e+00	3.2e-04	8.7e-03	-1.3e-02	1.4	5.4
hamming10-2.clq	524800	5120	523777	171		-2.33783e+00	-2.18000e+00	3.8e-04	5.3e-04	2.9e-02	13.4	48.4
hamming10-4.clq	524800	89600	523777	260		-1.93446e+01	-2.00798e+01	4.0e-04	1.5e-03	-1.8e-02	22.2	81.4
hamming6-2.clq	2080	192	2017	275	52	-1.89278e+00	-2.00537e+00	1.4e-03	1.3e-03	-2.3e-02	0.1	0.6
hamming6-4.clq	2080	1312	2017	477	104	-1.17603e+01	-1.20835e+01	1.3e-03	4.8e-03	-1.3e-02	0.3	1.1
hamming8-2.clq	32896	1024	32641	287		-2.17138e+00	-1.99072e+00	6.2e-04	1.3e-03	3.5e-02	0.9	3.4
hamming8-4.clq	32896	11776	32641	514		-1.56982e+01	-1.58373e+01	5.5e-04	5.1e-03	-4.3e-03	1.7	6.7
johnson16-2-4.clq	7260	1680	7141	505	278	-1.50610e+01	-1.51874e+01	3.3e-04	5.2e-03	-4.0e-03	0.5	1.9
johnson32-2-4.clq	123256	14880	122761	1053	420	-3.14188e+01	-3.11424e+01	5.7e-04	1.3e-03	4.3e-03	10.2	38.2
johnson8-2-4.clq	406	168	379	256	177	-6.89317e+00	-6.99401e+00	2.2e-03	5.2e-04	-6.8e-03	0.1	0.2
johnson8-4-4.clq	2485	560	2416	268	42	-5.03726e+00	-5.05317e+00	4.1e-04	5.1e-03	-1.4e-03	0.2	0.7
keller4.clq	14706	5100	14536	135		-1.51502e+01	-1.52925e+01	6.7e-04	4.7e-03	-4.5e-03	0.3	1.1
keller5.clq	301476	74710	300701	125		-3.03556e+01	-2.94901e+01	9.3e-05	5.5e-03	1.4e-02	4.7	17.2
keller6.clq	5649841	1026582	5646481	217		-6.28418e+01	-6.20364e+01	2.0e-04	2.5e-03	6.4e-03	770.8	2974.1
p-hat1000-1.clq	500500	377247	499501	108		-8.94539e+01	-8.74875e+01	1.7e-04	9.0e-03	1.1e-02	10.7	38.0
p-hat1000-2.clq	500500	254701	499501	128		-6.06763e+01	-5.89035e+01	1.3e-04	7.7e-03	1.5e-02	12.4	43.9
p-hat1000-3.clq	500500	127754	499501	35		-1.81843e+01	-1.81693e+01	3.6e-04	1.8e-03	4.0e-04	3.4	11.9
p-hat1500-1.clq	1125750	839327	1124251	118		-1.09173e+02	-1.06545e+02	1.3e-04	9.1e-03	1.2e-02	36.0	132.9
p-hat1500-2.clq	1125750	555290	1124251	138		-7.29887e+01	-7.06976e+01	1.0e-04	7.6e-03	1.6e-02	40.8	150.7
p-hat1500-3.clq	1125750	277006	1124251	35		-2.14808e+01	-2.14615e+01	3.0e-04	1.7e-03	4.4e-04	10.5	38.4
p-hat300-1.clq	45150	33917	44851	73		-4.40672e+01	-4.34590e+01	3.6e-04	8.5e-03	6.9e-03	0.6	2.2
p-hat300-2.clq	45150	22922	44851	82		-2.88892e+01	-2.82764e+01	3.0e-04	7.3e-03	1.1e-02	0.6	2.4
p-hat300-3.clq	45150	11460	44851	36		-1.04725e+01	-1.04197e+01	6.1e-04	2.2e-03	2.4e-03	0.3	1.1
p-hat500-1.clq	125250	93181	124751	84		-5.73066e+01	-5.62420e+01	2.6e-04	8.7e-03	9.3e-03	1.6	5.9
p-hat500-2.clq	125250	61804	124751	109		-3.89521e+01	-3.79550e+01	2.0e-04	7.5e-03	1.3e-02	2.0	7.6
p-hat500-3.clq	125250	30950	124751	35		-1.28753e+01	-1.28397e+01	5.0e-04	1.9e-03	1.3e-03	0.7	2.4
p-hat700-1.clq	245350	183651	244651	115		-7.34857e+01	-7.17007e+01	1.9e-04	9.0e-03	1.2e-02	4.6	16.6
p-hat700-2.clq	245350	122922	244651	151		-5.17165e+01	-5.01036e+01	1.4e-04	7.7e-03	1.6e-02	5.8	21.1
p-hat700-3.clq	245350	61640	244651	35		-1.52907e+01	-1.52649e+01	4.2e-04	1.8e-03	8.2e-04	1.4	5.0
san1000.clq	500500	249000	499501	234		-6.75641e+01	-6.55908e+01	3.8e-04	3.8e-03	1.5e-02	21.4	75.7
san200-0.7-1.clq	20100	5970	19901	59		-8.96491e+00	-9.08787e+00	6.9e-04	3.6e-03	-6.5e-03	0.2	0.8
san200-0.7-2.clq	20100	5970	19901	57		-1.16479e+01	-1.15035e+01	3.2e-04	5.8e-03	6.0e-03	0.2	0.8
san200-0.9-1.clq	20100	1990	19901	47		-4.07964e+00	-3.97131e+00	6.8e-04	2.3e-03	1.2e-02	0.2	0.6
san200-0.9-2.clq	20100	1990	19901	43		-4.36460e+00	-4.30649e+00	8.1e-04	1.4e-03	6.0e-03	0.1	0.6
san200-0.9-3.clq	20100	1990	19901	54		-5.02204e+00	-5.04936e+00	5.1e-04	2.2e-03	-2.5e-03	0.2	0.7
san400-0.5-1.clq	80200	39900	79801	58		-3.40707e+01	-3.43575e+01	4.4e-04	4.9e-03	-4.1e-03	0.8	2.8
san400-0.7-1.clq	80200	23940	79801	90		-1.24481e+01	-1.25538e+01	6.0e-04	2.2e-03	-4.1e-03	1.1	4.2
san400-0.7-2.clq	80200	23940	79801	40		-1.43344e+01	-1.42532e+01	4.8e-04	3.6e-03	2.7e-03	0.5	1.9
san400-0.7-3.clq	80200	23940	79801	421		-1.88422e+01	-1.85090e+01	4.1e-04	4.9e-03	8.7e-03	5.2	18.8
san400-0.9-1.clq	80200	7980	79801	48		-5.29960e+00	-5.18376e+00	4.9e-04	2.0e-03	1.0e-02	0.6	2.1
sanr200-0.7.clq	20100	6032	19901	36		-8.86193e+00	-8.84172e+00	7.2e-04	1.6e-03	1.1e-03	0.1	0.5
sanr200-0.9.clq	20100	2037	19901	39		-4.48831e+00	-4.47967e+00	7.8e-04	8.3e-04	8.7e-04	0.1	0.5
sanr400-0.5.clq	80200	39816	79801	40		-2.03422e+01	-2.01275e+01	4.5e-04	3.7e-03	5.2e-03	0.6	2.0
sanr400-0.7.clq	80200	23931	79801	37		-1.19300e+01	-1.20106e+01	5.1e-04	1.8e-03	-3.2e-03	0.5	1.8

Table 17: Algorithm-Performance with increased tolerance: Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=0.01

	$\frac{1}{2}(n_s^2 + n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	60		-7.72323e+00	-7.72035e+00	8.2e-05	1.1e-04	1.8e-04	0.2	0.9
brock200-2.clq	20100	10024	19901	58		-1.46075e+01	-1.45994e+01	6.3e-05	3.9e-04	2.7e-04	0.2	0.9
brock200-3.clq	20100	7852	19901	57		-1.11601e+01	-1.11634e+01	7.6e-05	1.5e-04	-1.4e-04	0.2	0.8
brock200-4.clq	20100	6811	19901	58		-9.82803e+00	-9.82457e+00	7.6e-05	1.2e-04	1.7e-04	0.2	0.8
brock400-1.clq	80200	20077	79801	59		-1.03568e+01	-1.03578e+01	5.8e-05	7.8e-05	-4.3e-05	0.8	2.8
brock400-2.clq	80200	20014	79801	59		-1.03751e+01	-1.03760e+01	6.0e-05	8.4e-05	-4.4e-05	0.8	2.8
brock400-3.clq	80200	20119	79801	59		-1.04055e+01	-1.04066e+01	6.0e-05	8.7e-05	-5.0e-05	0.8	2.8
brock400-4.clq	80200	20035	79801	59		-1.03990e+01	-1.03999e+01	5.9e-05	7.8e-05	-4.3e-05	0.8	2.8
brock800-1.clq	320400	112095	319601	58		-1.91506e+01	-1.91389e+01	3.8e-05	1.7e-04	3.0e-04	3.3	11.5
brock800-2.clq	320400	111434	319601	58		-1.90437e+01	-1.90330e+01	3.8e-05	1.6e-04	2.7e-04	3.2	11.5
brock800-3.clq	320400	112267	319601	58		-1.91863e+01	-1.91746e+01	3.8e-05	1.7e-04	3.0e-04	3.2	11.5
brock800-4.clq	320400	111957	319601	58		-1.91012e+01	-1.90896e+01	3.8e-05	1.6e-04	3.0e-04	3.2	11.5
c-fat200-1.clq	20100	18366	19901	465	167	-1.85228e+01	-1.84834e+01	7.3e-05	6.5e-04	1.0e-03	1.0	3.8
c-fat200-2.clq	20100	16665	19901	385	201	-9.02601e+00	-8.98814e+00	5.9e-05	8.0e-04	2.0e-03	0.7	2.7
c-fat200-5.clq	20100	11427	19901	181	126	-3.34412e+00	-3.31843e+00	8.8e-05	1.9e-04	3.4e-03	0.3	1.2
c-fat500-1.clq	125250	120291	124751	811	77	-3.99087e+01	-3.99533e+01	3.7e-05	8.9e-04	-5.5e-04	11.0	41.5
c-fat500-10.clq	125250	78123	124751	200	132	-4.03372e+00	-4.00307e+00	4.8e-05	5.1e-04	3.4e-03	1.8	6.6
c-fat500-2.clq	125250	115611	124751	533	123	-2.00485e+01	-2.00324e+01	3.0e-05	9.6e-04	3.9e-04	6.5	24.3
c-fat500-5.clq	125250	101559	124751	402	237	-8.05744e+00	-8.00454e+00	5.6e-05	3.5e-04	3.1e-03	3.9	14.3
hamming10-2.clq	524800	5120	523777	3321	889	-1.99582e+00	-1.99859e+00	3.9e-06	1.1e-04	-5.6e-04	222.4	768.3
hamming10-4.clq	524800	89600	523777	742	147	-2.00035e+01	-2.00277e+01	2.1e-06	4.7e-04	-5.9e-04	57.5	204.4
hamming6-2.clq	2080	192	2017	661	397	-2.01153e+00	-1.99989e+00	1.6e-04	2.7e-05	2.3e-03	0.3	0.9
hamming6-4.clq	2080	1312	2017	819	437	-1.20063e+01	-1.20153e+01	3.4e-05	8.8e-04	-3.6e-04	0.4	1.7
hamming8-2.clq	32896	1024	32641	1448	646	-2.00273e+00	-1.99860e+00	9.9e-06	2.0e-04	8.3e-04	3.7	14.4
hamming8-4.clq	32896	11776	32641	905	120	-1.59803e+01	-1.60198e+01	3.6e-05	6.2e-04	-1.2e-03	2.9	11.7
johnson16-2-4.clq	7260	1680	7141	796	566	-1.50187e+01	-1.49898e+01	1.0e-04	2.9e-04	9.3e-04	0.7	2.7
johnson32-2-4.clq	123256	14880	122761	1633	994	-3.10334e+01	-3.10280e+01	4.5e-05	2.6e-04	8.5e-05	14.0	51.3
johnson8-2-4.clq	406	168	379	406	326	-7.00390e+00	-7.00779e+00	8.0e-05	6.7e-04	-2.6e-04	0.1	0.3
johnson8-4-4.clq	2485	560	2416	447	219	-4.98674e+00	-4.99925e+00	1.5e-04	7.2e-05	-1.1e-03	0.2	0.9
keller4.clq	14706	5100	14536	297	13	-1.49979e+01	-1.50491e+01	1.1e-05	6.5e-04	-1.6e-03	0.5	2.1
keller5.clq	301476	74710	300701	732	20	-3.09854e+01	-3.11382e+01	1.6e-05	5.3e-04	-2.4e-03	26.6	96.8
keller6.clq	5649841	1026582	5646481	2661		-6.29704e+01	-6.32831e+01	7.7e-06	4.6e-04	-2.5e-03	9304.9	35914.7
p-hat1000-1.clq	500500	377247	499501	351		-8.94998e+01	-8.93301e+01	2.7e-05	7.7e-04	9.4e-04	34.8	124.3
p-hat1000-2.clq	500500	254701	499501	489		-6.08771e+01	-6.07203e+01	2.2e-05	7.1e-04	1.3e-03	46.6	166.7
p-hat1000-3.clq	500500	127754	499501	94		-1.81472e+01	-1.81391e+01	4.1e-05	6.7e-05	2.2e-04	8.7	31.0
p-hat1500-1.clq	1125750	839327	1124251	393		-1.09290e+02	-1.09058e+02	2.1e-05	7.9e-04	1.1e-03	119.5	441.6
p-hat1500-2.clq	1125750	555290	1124251	553		-7.33235e+01	-7.30980e+01	1.6e-05	7.2e-04	1.5e-03	162.8	601.1
p-hat1500-3.clq	1125750	277006	1124251	97		-2.14263e+01	-2.14169e+01	3.4e-05	6.6e-05	2.1e-04	28.0	103.2
p-hat300-1.clq	45150	33917	44851	223		-4.40798e+01	-4.40251e+01	5.6e-05	6.7e-04	6.1e-04	1.7	6.6
p-hat300-2.clq	45150	22922	44851	313		-2.89105e+01	-2.88583e+01	4.8e-05	6.3e-04	8.9e-04	2.3	9.0
p-hat300-3.clq	45150	11460	44851	106		-1.04329e+01	-1.04276e+01	7.4e-05	6.9e-05	2.4e-04	0.8	3.0
p-hat500-1.clq	125250	93181	124751	279		-5.72873e+01	-5.71989e+01	3.8e-05	7.5e-04	7.7e-04	5.2	19.7
p-hat500-2.clq	125250	61804	124751	441		-3.90847e+01	-3.89824e+01	3.1e-05	6.9e-04	1.3e-03	7.9	30.5
p-hat500-3.clq	125250	30950	124751	101		-1.28343e+01	-1.28285e+01	5.9e-05	6.7e-05	2.2e-04	1.8	6.8
p-hat700-1.clq	245350	183651	244651	347		-7.35181e+01	-7.33778e+01	3.4e-05	7.3e-04	9.5e-04	13.7	49.9
p-hat700-2.clq	245350	122922	244651	628		-5.20142e+01	-5.18501e+01	2.3e-05	7.3e-04	1.6e-03	23.8	86.8
p-hat700-3.clq	245350	61640	244651	98		-1.52497e+01	-1.52428e+01	4.9e-05	6.5e-05	2.2e-04	3.6	13.3
san1000.clq	500500	249000	499501	1579		-6.69360e+01	-6.67701e+01	2.9e-05	6.0e-04	1.2e-03	141.2	498.5
san200-0.7-1.clq	20100	5970	19901	147		-9.02368e+00	-9.01168e+00	5.2e-05	5.1e-04	6.3e-04	0.5	2.0
san200-0.7-2.clq	20100	5970	19901	836	69	-1.20122e+01	-1.19541e+01	6.7e-05	4.4e-04	2.3e-03	2.4	9.6
san200-0.9-1.clq	20100	1990	19901	260		-4.02451e+00	-4.01344e+00	6.9e-05	2.4e-04	1.2e-03	0.8	3.2
san200-0.9-2.clq	20100	1990	19901	185		-4.30746e+00	-4.29957e+00	7.8e-05	1.9e-04	8.2e-04	0.6	2.3
san200-0.9-3.clq	20100	1990	19901	85		-5.00125e+00	-4.99701e+00	8.2e-05	1.3e-04	3.9e-04	0.3	1.0
san400-0.5-1.clq	80200	39900	79801	104		-3.42110e+01	-3.41988e+01	4.2e-05	6.1e-04	1.7e-04	1.4	5.0
san400-0.7-1.clq	80200	23940	79801	170		-1.26314e+01	-1.25988e+01	5.0e-05	3.9e-04	1.2e-03	2.1	7.8
san400-0.7-2.clq	80200	23940	79801	8889	1358	-1.49808e+01	-1.48717e+01	6.1e-05	2.6e-04	3.5e-03	93.3	353.6
san400-0.7-3.clq	80200	23940	79801	2776	540	-1.89815e+01	-1.88894e+01	4.6e-05	4.6e-04	2.4e-03	30.6	110.1
san400-0.9-1.clq	80200	7980	79801	220		-5.21509e+00	-5.20029e+00	6.5e-05	8.1e-05	1.3e-03	2.6	9.6
sanr200-0.7.clq	20100	6032	19901	58		-8.83203e+00	-8.82948e+00	8.3e-05	1.1e-04	1.4e-04	0.2	0.8
sanr200-0.9.clq	20100	2037	19901	107		-4.48027e+00	-4.47723e+00	9.0e-05	6.3e-05	3.1e-04	0.3	1.3
sanr400-0.5.clq	80200	39816	79801	59		-2.01823e+01	-2.01661e+01	4.6e-05	3.2e-04	3.9e-04	0.8	2.9
sanr400-0.7.clq	80200	23931	79801	58		-1.20015e+01	-1.19955e+01	5.7e-05	8.7e-05	2.4e-04	0.8	2.8

Table 18: Algorithm-Performance with increased tolerance: Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=0.001

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	113	13	-7.71988e+00	-7.71970e+00	8.8e-06	8.9e-06	1.1e-05	0.4	1.6
brock200-2.clq	20100	10024	19901	79	9	-1.46060e+01	-1.46052e+01	7.4e-06	2.6e-05	2.6e-05	0.3	1.1
brock200-3.clq	20100	7852	19901	84	13	-1.11620e+01	-1.11619e+01	7.9e-06	1.7e-05	7.4e-06	0.3	1.2
brock200-4.clq	20100	6811	19901	88		-9.82492e+00	-9.82471e+00	8.3e-06	1.7e-05	1.0e-05	0.3	1.2
brock400-1.clq	80200	20077	79801	93		-1.03577e+01	-1.03575e+01	6.1e-06	1.1e-05	1.2e-05	1.2	4.3
brock400-2.clq	80200	20014	79801	96		-1.03759e+01	-1.03755e+01	6.0e-06	1.0e-05	1.9e-05	1.3	4.5
brock400-3.clq	80200	20119	79801	95		-1.04063e+01	-1.04060e+01	6.3e-06	1.1e-05	1.4e-05	1.2	4.5
brock400-4.clq	80200	20035	79801	94	7	-1.03999e+01	-1.03995e+01	6.0e-06	1.0e-05	1.8e-05	1.2	4.3
brock800-1.clq	320400	112095	319601	81		-1.91426e+01	-1.91431e+01	4.0e-06	1.0e-05	-1.2e-05	4.5	16.0
brock800-2.clq	320400	111434	319601	80		-1.90372e+01	-1.90366e+01	4.3e-06	1.1e-05	1.4e-05	4.4	15.7
brock800-3.clq	320400	112267	319601	81		-1.91781e+01	-1.91785e+01	4.0e-06	1.0e-05	-1.0e-05	4.5	16.0
brock800-4.clq	320400	111957	319601	80		-1.90939e+01	-1.90933e+01	4.3e-06	1.0e-05	1.3e-05	4.4	15.8
c-fat200-1.clq	20100	18366	19901	735	434	-1.84710e+01	-1.84686e+01	6.8e-06	7.4e-05	6.3e-05	1.3	5.3
c-fat200-2.clq	20100	16665	19901	670	484	-8.99780e+00	-8.99872e+00	5.0e-06	8.7e-05	-4.8e-05	1.1	4.3
c-fat200-5.clq	20100	11427	19901	279	223	-3.31605e+00	-3.31794e+00	5.4e-06	6.7e-05	-2.5e-04	0.4	1.7
c-fat500-1.clq	125250	120291	124751	1727	536	-4.00052e+01	-4.00072e+01	2.2e-06	1.0e-04	-2.5e-05	21.5	80.0
c-fat500-10.clq	125250	78123	124751	343	274	-3.99703e+00	-4.00037e+00	4.2e-06	6.2e-05	-3.7e-04	2.8	10.0
c-fat500-2.clq	125250	115611	124751	1111	647	-2.00031e+01	-2.00035e+01	1.9e-06	1.0e-04	-1.2e-05	11.4	41.1
c-fat500-5.clq	125250	101559	124751	688	520	-7.99417e+00	-8.00041e+00	5.7e-06	3.2e-05	-3.7e-04	6.0	21.5
hamming10-2.clq	524800	5120	523777	7041	4167	-1.99664e+00	-1.99990e+00	3.2e-06	7.7e-06	-6.5e-04	374.6	1172.2
hamming10-4.clq	524800	89600	523777	1229	615	-1.99951e+01	-2.00020e+01	3.0e-06	3.3e-05	-1.7e-04	83.5	280.2
hamming6-2.clq	2080	192	2017	1046	778	-1.99887e+00	-2.00003e+00	1.5e-05	8.4e-06	-2.3e-04	0.4	1.0
hamming6-4.clq	2080	1312	2017	1162	777	-1.20023e+01	-1.20009e+01	1.3e-05	5.3e-05	5.5e-05	0.6	2.3
hamming8-2.clq	32896	1024	32641	2609	1794	-1.99869e+00	-1.99988e+00	4.8e-06	1.7e-05	-2.4e-04	5.5	21.8
hamming8-4.clq	32896	11776	32641	1297	504	-1.60038e+01	-1.60012e+01	7.0e-06	3.6e-05	8.2e-05	4.0	16.0
johnson16-2-4.clq	7260	1680	7141	1087	854	-1.49984e+01	-1.49987e+01	8.6e-06	3.7e-05	-8.5e-06	0.9	3.5
johnson32-2-4.clq	123256	14880	122761	2214	1569	-3.10003e+01	-3.10042e+01	4.5e-07	4.0e-05	-6.1e-05	18.0	65.5
johnson8-2-4.clq	406	168	379	557	475	-7.00107e+00	-6.99998e+00	2.2e-05	2.1e-06	7.3e-05	0.2	0.4
johnson8-4-4.clq	2485	560	2416	627	397	-5.00127e+00	-4.99983e+00	1.4e-05	1.6e-05	1.3e-04	0.3	1.3
keller4.clq	14706	5100	14536	465	146	-1.49994e+01	-1.49954e+01	2.7e-06	6.2e-05	1.3e-04	0.8	3.1
keller5.clq	301476	74710	300701	1044	259	-3.09983e+01	-3.10137e+01	1.8e-06	5.2e-05	-2.4e-04	34.1	120.5
keller6.clq	5649841	1026582	5646481	3215	190	-6.29970e+01	-6.29608e+01	8.1e-07	4.6e-05	2.8e-04	10711.9	41093.7
p-hat1000-1.clq	500500	377247	499501	885		-8.94803e+01	-8.94782e+01	4.3e-06	5.7e-06	1.1e-05	88.9	315.3
p-hat1000-2.clq	500500	254701	499501	1812		-6.08660e+01	-6.08632e+01	4.3e-06	1.8e-06	2.2e-05	174.1	622.2
p-hat1000-3.clq	500500	127754	499501	274		-1.81372e+01	-1.81366e+01	4.3e-06	3.2e-06	1.4e-05	24.8	87.9
p-hat1500-1.clq	1125750	839327	1124251	968		-1.09267e+02	-1.09264e+02	3.5e-06	8.3e-06	1.6e-05	296.5	1096.5
p-hat1500-2.clq	1125750	555290	1124251	1863		-7.33130e+01	-7.33090e+01	3.6e-06	4.1e-06	2.7e-05	550.3	2030.8
p-hat1500-3.clq	1125750	277006	1124251	286		-2.14145e+01	-2.14139e+01	3.5e-06	3.0e-06	1.4e-05	81.4	298.7
p-hat300-1.clq	45150	33917	44851	654		-4.40689e+01	-4.40683e+01	7.7e-06	4.6e-06	7.0e-06	5.0	19.3
p-hat300-2.clq	45150	22922	44851	1343		-2.89023e+01	-2.89012e+01	7.7e-06	1.9e-06	2.0e-05	10.0	38.5
p-hat300-3.clq	45150	11460	44851	333		-1.04265e+01	-1.04261e+01	7.6e-06	2.9e-06	1.5e-05	2.3	9.2
p-hat500-1.clq	125250	93181	124751	725		-5.72753e+01	-5.72740e+01	6.0e-06	5.7e-06	1.1e-05	13.4	51.3
p-hat500-2.clq	125250	61804	124751	1640		-3.90771e+01	-3.90754e+01	6.0e-06	1.3e-06	2.1e-05	29.5	113.4
p-hat500-3.clq	125250	30950	124751	308		-1.28269e+01	-1.28265e+01	6.0e-06	2.8e-06	1.5e-05	5.3	20.2
p-hat700-1.clq	245350	183651	244651	1006		-7.34989e+01	-7.34978e+01	5.1e-06	4.1e-06	7.6e-06	39.9	144.8
p-hat700-2.clq	245350	122922	244651	2070		-5.20093e+01	-5.20064e+01	5.1e-06	9.2e-06	2.8e-05	78.8	287.1
p-hat700-3.clq	245350	61640	244651	311		-1.52413e+01	-1.52408e+01	5.1e-06	3.5e-06	1.5e-05	11.3	41.0
san1000.clq	500500	249000	499501	4503	1035	-6.70090e+01	-6.70092e+01	4.1e-06	2.4e-05	-1.4e-06	351.7	1209.7
san200-0.7-1.clq	20100	5970	19901	1641	542	-9.01978e+00	-9.01834e+00	3.7e-07	6.3e-05	7.6e-05	4.6	18.4
san200-0.7-2.clq	20100	5970	19901	1561	651	-1.20017e+01	-1.19991e+01	9.2e-06	1.0e-05	1.0e-04	3.7	14.7
san200-0.9-1.clq	20100	1990	19901	3087		-4.01900e+00	-4.01770e+00	1.5e-06	3.6e-05	1.4e-04	9.2	36.8
san200-0.9-2.clq	20100	1990	19901	502	24	-4.30186e+00	-4.30127e+00	8.7e-06	1.2e-05	6.2e-05	1.5	6.0
san200-0.9-3.clq	20100	1990	19901	266		-5.00014e+00	-4.99998e+00	9.2e-06	3.8e-06	1.4e-05	0.8	3.1
san400-0.5-1.clq	80200	39900	79801	244	14	-3.42062e+01	-3.42050e+01	2.5e-06	7.4e-05	1.8e-05	3.2	11.4
san400-0.7-1.clq	80200	23940	79801	303		-1.26165e+01	-1.26146e+01	6.5e-06	8.5e-06	7.2e-05	3.8	13.9
san400-0.7-2.clq	80200	23940	79801	23117	12512	-1.50008e+01	-1.50283e+01	2.5e-06	5.8e-05	-8.9e-04	179.0	667.6
san400-0.7-3.clq	80200	23940	79801	6357	3517	-1.90009e+01	-1.89856e+01	2.2e-06	5.9e-05	3.9e-04	51.8	186.2
san400-0.9-1.clq	80200	7980	79801	500		-5.20161e+00	-5.20003e+00	6.4e-06	1.0e-05	1.4e-04	6.0	21.7
sanr200-0.7.clq	20100	6032	19901	104		-8.82941e+00	-8.82925e+00	9.0e-06	8.1e-06	8.4e-06	0.4	1.5
sanr200-0.9.clq	20100	2037	19901	264		-4.47681e+00	-4.47658e+00	9.2e-06	3.7e-06	2.3e-05	0.8	3.3
sanr400-0.5.clq	80200	39816	79801	79	2	-2.01710e+01	-2.01701e+01	5.2e-06	2.3e-05	2.2e-05	1.1	3.8
sanr400-0.7.clq	80200	23931	79801	86	7	-1.19955e+01	-1.19950e+01	6.2e-06	1.1e-05	1.9e-05	1.1	3.9

Table 19: Algorithm-Performance with increased tolerance: Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=0.0001

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	372	231	-7.71969e+00	-7.71968e+00	9.2e-07	5.3e-07	4.3e-07	1.8	4.4
brock200-2.clq	20100	10024	19901	106	34	-1.46054e+01	-1.46054e+01	7.9e-07	2.1e-06	4.1e-07	0.4	1.6
brock200-3.clq	20100	7852	19901	123	48	-1.11618e+01	-1.11618e+01	8.2e-07	1.6e-06	5.4e-07	0.4	1.7
brock200-4.clq	20100	6811	19901	133	18	-9.82465e+00	-9.82463e+00	8.6e-07	1.4e-06	7.3e-07	0.5	2.0
brock400-1.clq	80200	20077	79801	142	5	-1.03575e+01	-1.03575e+01	6.1e-07	1.0e-06	1.3e-06	2.1	7.1
brock400-2.clq	80200	20014	79801	147		-1.03755e+01	-1.03755e+01	6.2e-07	1.0e-06	1.3e-06	2.2	7.4
brock400-3.clq	80200	20119	79801	146	26	-1.04060e+01	-1.04060e+01	6.2e-07	1.0e-06	1.4e-06	2.0	6.7
brock400-4.clq	80200	20035	79801	144	49	-1.03995e+01	-1.03995e+01	6.2e-07	1.0e-06	1.2e-06	1.8	5.9
brock800-1.clq	320400	112095	319601	110	10	-1.91431e+01	-1.91430e+01	4.3e-07	1.4e-06	1.6e-06	7.6	22.6
brock800-2.clq	320400	111434	319601	110	23	-1.90366e+01	-1.90365e+01	4.3e-07	1.3e-06	1.6e-06	7.0	20.9
brock800-3.clq	320400	112267	319601	111	3	-1.91785e+01	-1.91785e+01	4.2e-07	1.3e-06	4.0e-07	7.7	23.4
brock800-4.clq	320400	111957	319601	109		-1.90933e+01	-1.90932e+01	4.3e-07	1.3e-06	3.5e-07	7.6	23.0
c-fat200-1.clq	20100	18366	19901	2037	1723	-1.84666e+01	-1.84666e+01	8.7e-07	3.9e-06	1.3e-08	4.3	16.3
c-fat200-2.clq	20100	16665	19901	954	765	-8.99971e+00	-9.00011e+00	6.6e-07	7.2e-06	-2.1e-05	1.9	7.4
c-fat200-5.clq	20100	11427	19901	377	320	-3.31753e+00	-3.31764e+00	4.5e-07	7.3e-06	-1.4e-05	0.7	2.7
c-fat500-1.clq	125250	120291	124751	2766	1540	-3.99990e+01	-3.99995e+01	4.3e-07	7.9e-06	-5.6e-06	34.8	111.5
c-fat500-10.clq	125250	78123	124751	486	416	-3.99977e+00	-3.99996e+00	3.2e-07	7.4e-06	-2.0e-05	4.3	12.0
c-fat500-2.clq	125250	115611	124751	1689	1219	-2.00001e+01	-2.00004e+01	7.6e-08	1.1e-05	-6.2e-06	18.0	54.3
c-fat500-5.clq	125250	101559	124751	975	804	-8.00056e+00	-7.99995e+00	5.5e-07	4.1e-06	3.6e-05	9.2	26.3
hamming10-2.clq	524800	5120	523777	10762	7851	-1.99955e+00	-2.00000e+00	4.2e-07	1.7e-07	-8.9e-05	574.2	1608.7
hamming10-4.clq	524800	89600	523777	1715	1097	-1.99994e+01	-2.00001e+01	3.9e-07	1.8e-06	-1.8e-05	118.5	367.1
hamming6-2.clq	2080	192	2017	1432	1160	-2.00012e+00	-2.00000e+00	1.6e-06	2.4e-07	2.3e-05	1.1	2.0
hamming6-4.clq	2080	1312	2017	1505	1117	-1.20003e+01	-1.20000e+01	1.5e-06	7.6e-07	1.2e-05	1.2	5.0
hamming8-2.clq	32896	1024	32641	3770	2944	-1.99978e+00	-2.00000e+00	8.1e-07	4.6e-07	-4.4e-05	9.4	34.4
hamming8-4.clq	32896	11776	32641	1689	892	-1.60003e+01	-1.59998e+01	5.0e-07	5.5e-06	1.4e-05	5.8	21.4
johnson16-2-4.clq	7260	1680	7141	1379	1143	-1.49998e+01	-1.50001e+01	1.0e-06	2.8e-06	-9.3e-06	1.6	6.4
johnson32-2-4.clq	123256	14880	122761	2795	2145	-3.09997e+01	-3.10003e+01	3.9e-07	3.1e-06	-9.7e-06	24.0	72.8
johnson8-2-4.clq	406	168	379	707	624	-6.99995e+00	-6.99993e+00	1.0e-06	6.4e-06	1.7e-06	0.5	0.8
johnson8-4-4.clq	2485	560	2416	806	574	-4.99998e+00	-5.00006e+00	1.7e-07	5.3e-06	-6.4e-06	0.7	2.7
keller4.clq	14706	5100	14536	629	308	-1.50002e+01	-1.49998e+01	9.4e-07	2.3e-06	1.3e-05	1.2	4.7
keller5.clq	301476	74710	300701	1497	568	-3.10002e+01	-3.10013e+01	1.8e-07	5.2e-06	-1.8e-05	60.3	183.8
keller6.clq	5649841	1026582	5646481	3780	707	-6.29997e+01	-6.30039e+01	7.9e-08	4.6e-06	-3.3e-05	10900.4	41793.6
p-hat1000-1.clq	500500	377247	499501	2451		-8.94782e+01	-8.94782e+01	4.3e-07	1.4e-07	3.8e-07	301.1	930.2
p-hat1000-2.clq	500500	254701	499501	13309		-6.08627e+01	-6.08624e+01	4.3e-07	2.2e-08	2.0e-06	1570.0	4896.9
p-hat1000-3.clq	500500	127754	499501	871		-1.81366e+01	-1.81366e+01	4.3e-07	1.2e-07	5.8e-07	92.1	293.5
p-hat1500-1.clq	1125750	839327	1124251	2475		-1.09265e+02	-1.09265e+02	3.6e-07	1.7e-07	3.9e-07	867.9	2833.8
p-hat1500-2.clq	1125750	555290	1124251	13486		-7.33090e+01	-7.33087e+01	3.6e-07	2.8e-08	2.0e-06	4587.0	15037.8
p-hat1500-3.clq	1125750	277006	1124251	874		-2.14139e+01	-2.14138e+01	3.6e-07	1.3e-07	5.8e-07	266.7	886.6
p-hat300-1.clq	45150	33917	44851	2068	74	-4.40680e+01	-4.40679e+01	7.7e-07	1.8e-07	5.0e-07	18.2	66.1
p-hat300-2.clq	45150	22922	44851	7999		-2.89009e+01	-2.89008e+01	7.7e-07	3.4e-08	1.6e-06	68.6	254.9
p-hat300-3.clq	45150	11460	44851	928		-1.04261e+01	-1.04261e+01	7.7e-07	1.5e-07	7.1e-07	7.4	28.1
p-hat500-1.clq	125250	93181	124751	2066	424	-5.72741e+01	-5.72741e+01	6.0e-07	1.3e-07	4.0e-07	45.5	144.7
p-hat500-2.clq	125250	61804	124751	11773		-3.90750e+01	-3.90749e+01	6.1e-07	3.2e-08	1.9e-06	278.7	909.1
p-hat500-3.clq	125250	30950	124751	849		-1.28265e+01	-1.28265e+01	6.0e-07	1.2e-07	6.6e-07	18.1	60.7
p-hat700-1.clq	245350	183651	244651	3165		-7.34972e+01	-7.34971e+01	5.1e-07	1.0e-07	5.2e-07	162.4	497.8
p-hat700-2.clq	245350	122922	244651	15666		-5.20065e+01	-5.20063e+01	5.1e-07	5.3e-08	2.0e-06	771.6	2400.1
p-hat700-3.clq	245350	61640	244651	1046		-1.52408e+01	-1.52407e+01	5.1e-07	1.4e-07	7.8e-07	46.5	147.8
san1000.clq	500500	249000	499501	7426	3746	-6.69991e+01	-6.70012e+01	4.0e-07	3.0e-06	-1.5e-05	558.7	1623.1
san200-0.7-1.clq	20100	5970	19901	6438	4711	-9.01999e+00	-9.01980e+00	2.5e-08	6.3e-06	9.7e-06	16.3	63.3
san200-0.7-2.clq	20100	5970	19901	2286	1369	-1.20001e+01	-1.20003e+01	8.2e-07	2.9e-06	-6.2e-06	5.7	22.3
san200-0.9-1.clq	20100	1990	19901	10196		-4.01912e+00	-4.01897e+00	2.4e-07	3.5e-06	1.7e-05	36.2	144.6
san200-0.9-2.clq	20100	1990	19901	924	344	-4.30144e+00	-4.30139e+00	9.2e-07	6.1e-07	5.0e-06	2.8	10.9
san200-0.9-3.clq	20100	1990	19901	1132		-5.00001e+00	-5.00000e+00	9.3e-07	2.0e-07	1.2e-06	3.9	15.7
san400-0.5-1.clq	80200	39900	79801	420	148	-3.42059e+01	-3.42058e+01	1.8e-07	7.6e-06	1.4e-06	5.2	17.3
san400-0.7-1.clq	80200	23940	79801	457	99	-1.26149e+01	-1.26148e+01	6.5e-07	8.0e-07	6.3e-06	5.7	19.4
san400-0.7-2.clq	80200	23940	79801	37346	26510	-1.50001e+01	-1.49972e+01	2.9e-07	5.7e-06	9.2e-05	275.4	872.3
san400-0.7-3.clq	80200	23940	79801	9939	7064	-1.90003e+01	-1.90000e+01	6.7e-07	1.1e-07	6.2e-06	78.4	241.2
san400-0.9-1.clq	80200	7980	79801	787	48	-5.20025e+00	-5.20010e+00	6.0e-07	1.5e-06	1.3e-05	10.0	35.1
sanr200-0.7.clq	20100	6032	19901	210	2	-8.82923e+00	-8.82922e+00	9.1e-07	3.6e-07	2.9e-07	0.9	3.4
sanr200-0.9.clq	20100	2037	19901	623		-4.47656e+00	-4.47656e+00	9.3e-07	9.8e-08	6.0e-07	2.3	9.2
sanr400-0.5.clq	80200	39816	79801	103	21	-2.01700e+01	-2.01700e+01	5.9e-07	1.9e-06	-3.4e-07	1.5	4.8
sanr400-0.7.clq	80200	23931	79801	129	44	-1.19950e+01	-1.19950e+01	6.1e-07	1.2e-06	1.1e-06	1.6	5.1

Table 20: Algorithm-Performance with increased tolerance: Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=0-maxtime=Inf-maxit=Inf-tol=1e-05

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	351	86	-7.71968e+00	-7.71968e+00	5.2e-08	2.7e-07	-3.6e-09	2.4	9.5
brock200-2.clq	20100	10024	19901	106	48	-1.46054e+01	-1.46054e+01	6.3e-08	4.3e-07	2.5e-09	0.7	2.6
brock200-3.clq	20100	7852	19901	112	51	-1.11617e+01	-1.11617e+01	6.0e-08	4.1e-07	-5.3e-09	0.7	2.7
brock200-4.clq	20100	6811	19901	118	34	-9.82463e+00	-9.82463e+00	6.0e-08	3.2e-07	2.9e-09	0.8	3.0
brock400-1.clq	80200	20077	79801	95	18	-1.03575e+01	-1.03575e+01	4.4e-08	2.9e-07	-3.7e-09	1.8	6.7
brock400-2.clq	80200	20014	79801	98	1	-1.03755e+01	-1.03755e+01	4.5e-08	2.9e-07	1.1e-08	2.0	7.4
brock400-3.clq	80200	20119	79801	99	30	-1.04060e+01	-1.04060e+01	4.6e-08	3.2e-07	-7.8e-09	1.8	6.7
brock400-4.clq	80200	20035	79801	100	40	-1.03994e+01	-1.03994e+01	4.0e-08	2.9e-07	6.4e-09	1.7	6.6
brock800-1.clq	320400	112095	319601	87	22	-1.91430e+01	-1.91430e+01	2.9e-08	3.2e-07	5.7e-09	6.8	23.9
brock800-2.clq	320400	111434	319601	86	27	-1.90365e+01	-1.90365e+01	3.2e-08	3.6e-07	-3.5e-09	6.6	23.2
brock800-3.clq	320400	112267	319601	87	16	-1.91785e+01	-1.91785e+01	3.2e-08	3.5e-07	4.9e-09	6.9	24.6
brock800-4.clq	320400	111957	319601	86	13	-1.90932e+01	-1.90932e+01	3.1e-08	3.4e-07	-4.1e-09	6.9	24.7
c-fat200-1.clq	20100	18366	19901	100000	4930	-1.84661e+01	-1.84750e+01	5.7e-04	5.3e-03	-2.3e-04	683.8	2734.4
c-fat200-2.clq	20100	16665	19901	142	118	-9.00000e+00	-9.00000e+00	4.8e-08	5.6e-07	2.8e-07	0.7	2.8
c-fat200-5.clq	20100	11427	19901	98	86	-3.31767e+00	-3.31767e+00	4.5e-08	3.9e-07	6.1e-08	0.5	1.9
c-fat500-1.clq	125250	120291	124751	182	26	-4.00000e+01	-4.00000e+01	3.3e-08	6.3e-07	9.8e-08	5.0	19.2
c-fat500-10.clq	125250	78123	124751	145	132	-4.00000e+00	-4.00000e+00	3.0e-08	4.9e-07	-4.2e-07	2.9	10.8
c-fat500-2.clq	125250	115611	124751	79	35	-2.00000e+01	-2.00000e+01	2.8e-08	5.3e-07	1.8e-07	1.9	7.3
c-fat500-5.clq	125250	101559	124751	134	112	-8.00001e+00	-8.00000e+00	3.2e-08	7.3e-07	5.1e-07	2.9	10.9
hamming10-2.clq	524800	5120	523777	804	551	-2.00000e+00	-2.00000e+00	2.7e-08	7.2e-08	3.5e-08	81.5	260.1
hamming10-4.clq	524800	89600	523777	71	53	-2.00000e+01	-2.00000e+01	1.7e-08	1.0e-07	-1.5e-08	7.7	25.1
hamming6-2.clq	2080	192	2017	182	123	-2.00000e+00	-2.00000e+00	7.8e-08	2.5e-07	5.5e-09	0.3	1.3
hamming6-4.clq	2080	1312	2017	55	43	-1.20000e+01	-1.20000e+01	3.8e-08	2.3e-07	5.4e-09	0.1	0.4
hamming8-2.clq	32896	1024	32641	132	101	-2.00000e+00	-2.00000e+00	6.1e-08	1.0e-07	-2.0e-08	0.8	3.2
hamming8-4.clq	32896	11776	32641	51	38	-1.60000e+01	-1.60000e+01	1.7e-08	6.4e-08	-1.5e-07	0.3	1.3
johnson16-2-4.clq	7260	1680	7141	82000	81181	-1.50000e+01	-1.50000e+01	2.0e-08	3.1e-08	1.2e-09	187.6	746.8
johnson32-2-4.clq	123256	14880	122761	13	4	-3.10000e+01	-3.10000e+01	3.5e-09	4.2e-09	3.3e-08	0.2	0.9
johnson8-2-4.clq	406	168	379	13	5	-7.00000e+00	-7.00000e+00	5.2e-09	2.6e-09	-3.9e-09	0.0	0.0
johnson8-4-4.clq	2485	560	2416	27	14	-5.00000e+00	-5.00000e+00	2.3e-08	6.8e-08	-2.1e-07	0.0	0.2
keller4.clq	14706	5100	14536	84	15	-1.50000e+01	-1.50000e+01	3.3e-08	4.0e-07	3.7e-08	0.5	1.9
keller5.clq	301476	74710	300701	132	21	-3.10000e+01	-3.10000e+01	1.4e-08	3.0e-07	4.2e-08	8.1	29.0
keller6.clq	5649841	1026582	5646481	267	11	-6.30000e+01	-6.30000e+01	1.2e-08	2.7e-07	-3.1e-07	1057.9	3931.3
p-hat1000-1.clq	500500	377247	499501	9151	3877	-8.94781e+01	-8.94781e+01	3.3e-08	6.3e-07	2.2e-08	1292.7	4480.0
p-hat1000-2.clq	500500	254701	499501	100000		-6.08624e+01	-6.08624e+01	5.6e-08	1.0e-06	3.6e-09	14970.2	53006.5
p-hat1000-3.clq	500500	127754	499501	892		-1.81366e+01	-1.81366e+01	3.0e-08	3.4e-07	2.3e-08	122.4	430.2
p-hat1500-1.clq	1125750	839327	1124251	4129	1076	-1.09265e+02	-1.09265e+02	2.5e-08	6.1e-07	7.4e-09	1656.0	5828.8
p-hat1500-2.clq	1125750	555290	1124251	100000		-7.33086e+01	-7.33086e+01	4.7e-08	9.5e-07	3.5e-09	41950.3	150894.0
p-hat1500-3.clq	1125750	277006	1124251	1185		-2.14138e+01	-2.14138e+01	2.7e-08	3.5e-07	-8.4e-08	464.6	1664.4
p-hat300-1.clq	45150	33917	44851	4052		-4.40679e+01	-4.40679e+01	4.5e-08	5.6e-07	-1.3e-10	53.9	210.7
p-hat300-2.clq	45150	22922	44851	38240		-2.89008e+01	-2.89008e+01	5.9e-08	5.1e-07	5.8e-10	486.7	1910.0
p-hat300-3.clq	45150	11460	44851	1560		-1.04261e+01	-1.04261e+01	5.8e-08	3.6e-07	2.1e-09	18.8	74.2
p-hat500-1.clq	125250	93181	124751	4233	1228	-5.72741e+01	-5.72741e+01	4.6e-08	6.2e-07	2.7e-09	125.8	482.0
p-hat500-2.clq	125250	61804	124751	100000		-3.90748e+01	-3.90748e+01	6.2e-08	8.0e-07	1.7e-09	3032.9	11750.3
p-hat500-3.clq	125250	30950	124751	892		-1.28265e+01	-1.28265e+01	4.5e-08	3.6e-07	-3.6e-08	25.3	97.5
p-hat700-1.clq	245350	183651	244651	9003	1360	-7.34971e+01	-7.34971e+01	3.6e-08	5.8e-07	-3.9e-09	568.6	2070.8
p-hat700-2.clq	245350	122922	244651	100000		-5.20063e+01	-5.20063e+01	8.9e-08	1.3e-06	4.1e-09	6317.5	23221.2
p-hat700-3.clq	245350	61640	244651	2130		-1.52407e+01	-1.52407e+01	3.9e-08	3.6e-07	2.8e-08	123.1	450.2
san1000.clq	500500	249000	499501	268		-6.70000e+01	-6.70000e+01	2.7e-08	4.8e-07	1.0e-08	37.7	132.6
san200-0.7-1.clq	20100	5970	19901	2383	1430	-9.01999e+00	-9.01999e+00	5.6e-08	3.5e-07	2.0e-08	13.4	53.6
san200-0.7-2.clq	20100	5970	19901	141		-1.20000e+01	-1.20000e+01	4.6e-08	3.1e-07	-3.4e-11	0.9	3.6
san200-0.9-1.clq	20100	1990	19901	20281		-4.01912e+00	-4.01912e+00	6.4e-08	2.1e-07	3.9e-08	127.2	508.7
san200-0.9-2.clq	20100	1990	19901	588	255	-4.30140e+00	-4.30140e+00	6.6e-08	1.8e-07	1.1e-07	3.3	13.0
san200-0.9-3.clq	20100	1990	19901	3825		-5.00000e+00	-5.00000e+00	7.4e-08	2.2e-07	-1.2e-08	23.6	94.4
san400-0.5-1.clq	80200	39900	79801	326	136	-3.42059e+01	-3.42059e+01	4.1e-08	4.0e-07	-1.8e-09	6.0	22.6
san400-0.7-1.clq	80200	23940	79801	217	76	-1.26148e+01	-1.26148e+01	4.9e-08	3.2e-07	-5.0e-09	3.9	14.9
san400-0.7-2.clq	80200	23940	79801	343	28	-1.50000e+01	-1.50000e+01	3.0e-08	2.8e-07	-2.9e-09	6.5	25.0
san400-0.7-3.clq	80200	23940	79801	97	17	-1.90000e+01	-1.90000e+01	3.0e-08	4.3e-07	-3.0e-08	1.8	6.8
san400-0.9-1.clq	80200	7980	79801	472	120	-5.20011e+00	-5.20011e+00	4.8e-08	2.3e-07	-3.2e-08	8.1	31.0
sanr200-0.7.clq	20100	6032	19901	153	29	-8.82922e+00	-8.82922e+00	5.6e-08	3.1e-07	-6.5e-09	1.1	4.2
sanr200-0.9.clq	20100	2037	19901	442		-4.47656e+00	-4.47656e+00	6.7e-08	2.2e-07	-1.7e-07	2.8	11.2
sanr400-0.5.clq	80200	39816	79801	92	31	-2.01700e+01	-2.01700e+01	3.5e-08	4.2e-07	-1.1e-09	1.7	6.5
sanr400-0.7.clq	80200	23931	79801	93	38	-1.19950e+01	-1.19950e+01	4.3e-08	3.1e-07	2.2e-09	1.7	6.3

Table 21: Algorithm-Performance NCQNM(5): Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=5-maxtime=Inf-maxit=100000-tol=1e-06

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	280	30	-7.71968e+00	-7.71968e+00	5.7e-08	3.2e-07	-1.1e-08	2.3	9.0
brock200-2.clq	20100	10024	19901	102	45	-1.46054e+01	-1.46054e+01	5.3e-08	4.1e-07	4.3e-10	0.7	3.0
brock200-3.clq	20100	7852	19901	104	46	-1.11618e+01	-1.11618e+01	6.8e-08	4.0e-07	-1.3e-09	0.8	3.1
brock200-4.clq	20100	6811	19901	105	27	-9.82463e+00	-9.82463e+00	5.3e-08	3.6e-07	1.9e-09	0.8	3.2
brock400-1.clq	80200	20077	79801	93	17	-1.03575e+01	-1.03575e+01	4.1e-08	2.7e-07	-6.7e-10	2.1	7.8
brock400-2.clq	80200	20014	79801	95	2	-1.03755e+01	-1.03755e+01	4.0e-08	2.6e-07	8.5e-10	2.2	8.4
brock400-3.clq	80200	20119	79801	95	27	-1.04060e+01	-1.04060e+01	4.4e-08	2.9e-07	2.4e-09	2.1	7.8
brock400-4.clq	80200	20035	79801	95	36	-1.03994e+01	-1.03994e+01	4.2e-08	2.7e-07	6.8e-10	2.0	7.5
brock800-1.clq	320400	112095	319601	84	20	-1.91430e+01	-1.91430e+01	3.1e-08	3.3e-07	-1.3e-09	7.5	27.3
brock800-2.clq	320400	111434	319601	84	26	-1.90365e+01	-1.90365e+01	3.0e-08	3.2e-07	-2.2e-09	7.3	26.3
brock800-3.clq	320400	112267	319601	84	15	-1.91785e+01	-1.91785e+01	3.2e-08	3.4e-07	4.0e-10	7.6	27.7
brock800-4.clq	320400	111957	319601	83	12	-1.90932e+01	-1.90932e+01	3.4e-08	3.5e-07	2.0e-09	7.6	27.5
c-fat200-1.clq	20100	18366	19901	100000	25760	-1.84782e+01	-1.84779e+01	5.0e-04	5.7e-03	8.6e-06	802.4	3208.7
c-fat200-2.clq	20100	16665	19901	131	109	-9.00000e+00	-9.00000e+00	3.6e-08	4.9e-07	6.1e-08	0.8	3.3
c-fat200-5.clq	20100	11427	19901	78	66	-3.31767e+00	-3.31767e+00	2.8e-08	2.3e-07	1.1e-07	0.5	1.8
c-fat500-1.clq	125250	120291	124751	168	13	-4.00000e+01	-4.00000e+01	3.4e-08	5.9e-07	-3.0e-08	5.7	22.2
c-fat500-10.clq	125250	78123	124751	135	122	-4.00001e+00	-4.00000e+00	2.5e-08	4.1e-07	9.4e-07	3.4	13.2
c-fat500-2.clq	125250	115611	124751	65	25	-2.00000e+01	-2.00000e+01	3.7e-08	6.2e-07	-2.7e-08	2.0	7.6
c-fat500-5.clq	125250	101559	124751	63	42	-8.00000e+00	-8.00000e+00	2.8e-08	5.7e-07	-1.7e-07	1.7	6.6
hamming10-2.clq	524800	5120	523777	2379	1892	-2.00004e+00	-2.00000e+00	3.8e-08	3.9e-08	7.3e-06	268.2	879.1
hamming10-4.clq	524800	89600	523777	40762	40340	-2.00000e+01	-2.00000e+01	1.6e-08	2.0e-07	6.0e-08	4886.4	16058.9
hamming6-2.clq	2080	192	2017	73	52	-2.00000e+00	-2.00000e+00	3.7e-09	1.1e-08	1.2e-10	0.1	0.5
hamming6-4.clq	2080	1312	2017	17	5	-1.20000e+01	-1.20000e+01	6.3e-08	9.1e-08	-4.0e-07	0.0	0.1
hamming8-2.clq	32896	1024	32641	307	184	-2.00001e+00	-2.00000e+00	5.9e-08	1.3e-07	1.2e-06	2.7	10.5
hamming8-4.clq	32896	11776	32641	19	6	-1.60000e+01	-1.60000e+01	8.2e-09	1.6e-07	-9.2e-08	0.1	0.6
johnson16-2-4.clq	7260	1680	7141	17	9	-1.50000e+01	-1.50000e+01	1.4e-08	2.7e-08	-5.5e-08	0.1	0.2
johnson32-2-4.clq	123256	14880	122761	13	4	-3.10000e+01	-3.10000e+01	2.6e-09	2.6e-09	1.2e-08	0.3	1.0
johnson8-2-4.clq	406	168	379	13	5	-7.00000e+00	-7.00000e+00	5.2e-09	2.6e-09	-3.9e-09	0.0	0.1
johnson8-4-4.clq	2485	560	2416	18	5	-5.00000e+00	-5.00000e+00	6.1e-09	1.4e-09	1.4e-09	0.0	0.1
keller4.clq	14706	5100	14536	83	28	-1.50000e+01	-1.50000e+01	4.7e-08	3.4e-07	1.6e-08	0.5	2.2
keller5.clq	301476	74710	300701	128	29	-3.10000e+01	-3.10000e+01	2.9e-08	3.0e-07	1.3e-07	9.0	33.2
keller6.clq	5649841	1026582	5646481	322	60	-6.30000e+01	-6.29996e+01	9.2e-09	4.5e-07	3.7e-06	1234.0	4545.8
p-hat1000-1.clq	500500	377247	499501	13163	9189	-8.94782e+01	-8.94780e+01	3.0e-08	6.5e-07	1.2e-06	1999.8	7026.3
p-hat1000-2.clq	500500	254701	499501	81197		-6.08624e+01	-6.08623e+01	3.2e-08	5.0e-07	5.7e-07	13866.4	50103.5
p-hat1000-3.clq	500500	127754	499501	736		-1.81366e+01	-1.81366e+01	2.9e-08	3.8e-07	4.6e-08	114.6	411.9
p-hat1500-1.clq	1125750	839327	1124251	2589	453	-1.09265e+02	-1.09265e+02	2.2e-08	5.3e-07	-1.2e-08	1206.6	4344.0
p-hat1500-2.clq	1125750	555290	1124251	82479		-7.33087e+01	-7.33086e+01	2.4e-08	5.2e-07	5.3e-07	38391.9	139839.8
p-hat1500-3.clq	1125750	277006	1124251	854		-2.14138e+01	-2.14138e+01	2.5e-08	3.7e-07	-4.3e-08	371.6	1344.3
p-hat300-1.clq	45150	33917	44851	2561		-4.40680e+01	-4.40679e+01	6.2e-08	5.6e-07	4.9e-07	41.7	163.7
p-hat300-2.clq	45150	22922	44851	19128		-2.89008e+01	-2.89008e+01	5.6e-08	5.0e-07	3.0e-07	295.4	1163.3
p-hat300-3.clq	45150	11460	44851	1184		-1.04261e+01	-1.04261e+01	5.6e-08	3.6e-07	6.4e-08	17.2	68.1
p-hat500-1.clq	125250	93181	124751	2786	680	-5.72741e+01	-5.72741e+01	4.4e-08	4.7e-07	5.7e-07	100.0	388.2
p-hat500-2.clq	125250	61804	124751	67415		-3.90749e+01	-3.90748e+01	4.1e-08	5.3e-07	5.4e-07	2405.5	9391.5
p-hat500-3.clq	125250	30950	124751	615		-1.28265e+01	-1.28265e+01	4.0e-08	3.4e-07	7.4e-09	20.4	79.0
p-hat700-1.clq	245350	183651	244651	12376	6836	-7.34971e+01	-7.34969e+01	3.3e-08	7.3e-07	1.3e-06	829.7	3044.7
p-hat700-2.clq	245350	122922	244651	92600		-5.20063e+01	-5.20062e+01	4.0e-08	4.7e-07	7.4e-07	6704.1	25105.4
p-hat700-3.clq	245350	61640	244651	1527		-1.52407e+01	-1.52407e+01	3.6e-08	4.0e-07	6.1e-08	102.5	381.5
san1000.clq	500500	249000	499501	261		-6.70000e+01	-6.70000e+01	2.8e-08	4.9e-07	-7.2e-08	42.1	150.9
san200-0.7-1.clq	20100	5970	19901	2036	1131	-9.01999e+00	-9.01999e+00	5.4e-08	3.8e-07	-8.5e-08	14.4	57.3
san200-0.7-2.clq	20100	5970	19901	130		-1.20000e+01	-1.20000e+01	5.7e-08	4.1e-07	-3.6e-08	1.0	3.9
san200-0.9-1.clq	20100	1990	19901	15774		-4.01912e+00	-4.01912e+00	6.9e-08	2.4e-07	1.5e-07	117.8	471.3
san200-0.9-2.clq	20100	1990	19901	574	278	-4.30140e+00	-4.30140e+00	6.0e-08	2.0e-07	1.1e-07	3.9	15.5
san200-0.9-3.clq	20100	1990	19901	2937		-5.00000e+00	-5.00000e+00	7.4e-08	2.2e-07	-6.9e-10	21.5	86.0
san400-0.5-1.clq	80200	39900	79801	316	131	-3.42059e+01	-3.42059e+01	4.5e-08	5.4e-07	-2.5e-10	7.0	26.8
san400-0.7-1.clq	80200	23940	79801	190	67	-1.26148e+01	-1.26148e+01	4.5e-08	3.3e-07	4.4e-09	4.0	15.5
san400-0.7-2.clq	80200	23940	79801	325	61	-1.50000e+01	-1.50000e+01	3.2e-08	4.5e-07	2.0e-08	7.0	27.1
san400-0.7-3.clq	80200	23940	79801	107	25	-1.90000e+01	-1.90000e+01	3.7e-08	2.5e-07	-3.4e-08	2.3	8.8
san400-0.9-1.clq	80200	7980	79801	388	103	-5.20011e+00	-5.20011e+00	4.4e-08	2.0e-07	5.9e-08	8.0	30.6
sanr200-0.7.clq	20100	6032	19901	135	23	-8.82922e+00	-8.82922e+00	7.0e-08	3.3e-07	5.1e-09	1.1	4.2
sanr200-0.9.clq	20100	2037	19901	336	33	-4.47656e+00	-4.47656e+00	6.3e-08	2.5e-07	-1.1e-07	2.5	10.1
sanr400-0.5.clq	80200	39816	79801	91	31	-2.01700e+01	-2.01700e+01	3.9e-08	3.9e-07	2.8e-09	2.1	7.8
sanr400-0.7.clq	80200	23931	79801	90	35	-1.19950e+01	-1.19950e+01	4.3e-08	3.0e-07	1.8e-09	1.9	7.2

Table 22: Algorithm-Performance NCQNM(10): Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=10-maxtime=Inf-maxit=100000-tol=1e-06

	$\frac{1}{2}(n_s^2+n_s)$	n_l	m	**it**	**up**	$C \bullet X$	$b^T y$	R_P	R_D	gap	**time**	**cpu**
brock200-1.clq	20100	5066	19901	294	28	-7.71968e+00	-7.71969e+00	6.4e-08	3.8e-07	-4.1e-07	3.2	12.5
brock200-2.clq	20100	10024	19901	99	42	-1.46054e+01	-1.46054e+01	5.7e-08	4.1e-07	-4.5e-09	1.1	4.2
brock200-3.clq	20100	7852	19901	102	46	-1.11618e+01	-1.11618e+01	5.7e-08	4.1e-07	6.8e-08	0.9	3.7
brock200-4.clq	20100	6811	19901	103	28	-9.82463e+00	-9.82463e+00	4.8e-08	3.7e-07	-3.7e-08	1.0	4.1
brock400-1.clq	80200	20077	79801	92	16	-1.03575e+01	-1.03575e+01	5.5e-08	3.2e-07	3.7e-08	2.4	9.0
brock400-2.clq	80200	20014	79801	94	1	-1.03755e+01	-1.03755e+01	5.3e-08	3.2e-07	3.8e-08	2.6	9.9
brock400-3.clq	80200	20119	79801	94	27	-1.04060e+01	-1.04060e+01	5.3e-08	3.2e-07	4.3e-09	2.4	9.1
brock400-4.clq	80200	20035	79801	93	34	-1.03994e+01	-1.03994e+01	5.0e-08	3.2e-07	-3.5e-09	2.4	9.1
brock800-1.clq	320400	112095	319601	84	20	-1.91430e+01	-1.91430e+01	2.6e-08	3.3e-07	6.1e-10	8.6	31.4
brock800-2.clq	320400	111434	319601	83	25	-1.90365e+01	-1.90365e+01	3.3e-08	3.5e-07	-2.8e-09	8.3	30.3
brock800-3.clq	320400	112267	319601	84	15	-1.91785e+01	-1.91785e+01	3.1e-08	3.4e-07	7.7e-10	8.5	31.0
brock800-4.clq	320400	111957	319601	83	12	-1.90932e+01	-1.90932e+01	3.2e-08	3.4e-07	-2.9e-09	8.6	31.5
c-fat200-1.clq	20100	18366	19901	100000	42334	-1.84664e+01	-1.84704e+01	3.3e-04	3.9e-03	-1.1e-04	996.4	3984.6
c-fat200-2.clq	20100	16665	19901	135	114	-9.00000e+00	-9.00000e+00	3.6e-08	3.6e-07	-2.0e-07	1.2	4.6
c-fat200-5.clq	20100	11427	19901	331	316	-3.31767e+00	-3.31767e+00	8.8e-08	2.6e-07	9.0e-07	2.3	9.3
c-fat500-1.clq	125250	120291	124751	175	11	-4.00000e+01	-4.00000e+01	2.7e-08	5.4e-07	-1.3e-08	7.4	28.8
c-fat500-10.clq	125250	78123	124751	132	119	-4.00000e+00	-4.00000e+00	2.8e-08	2.9e-07	1.5e-07	4.1	15.9
c-fat500-2.clq	125250	115611	124751	65	26	-2.00000e+01	-2.00000e+01	2.1e-08	4.9e-07	7.6e-08	2.4	9.2
c-fat500-5.clq	125250	101559	124751	130	110	-8.00000e+00	-8.00000e+00	2.1e-08	6.0e-07	-2.0e-08	4.2	16.4
hamming10-2.clq	524800	5120	523777	192	141	-2.00000e+00	-2.00000e+00	4.6e-09	7.3e-09	-3.5e-07	24.9	84.4
hamming10-4.clq	524800	89600	523777	9101	8987	-2.00000e+01	-2.00000e+01	1.7e-08	1.7e-07	4.5e-07	1275.8	4339.1
hamming6-2.clq	2080	192	2017	30	9	-2.00000e+00	-2.00000e+00	5.2e-08	8.6e-09	1.4e-07	0.1	0.2
hamming6-4.clq	2080	1312	2017	17	5	-1.20000e+01	-1.20000e+01	4.7e-10	3.4e-10	1.7e-10	0.0	0.1
hamming8-2.clq	32896	1024	32641	10640	10462	-1.99999e+00	-2.00000e+00	3.6e-08	1.8e-08	-1.7e-06	85.9	342.1
hamming8-4.clq	32896	11776	32641	19	6	-1.60000e+01	-1.60000e+01	3.2e-10	2.3e-09	9.2e-10	0.2	0.6
johnson16-2-4.clq	7260	1680	7141	17	9	-1.50000e+01	-1.50000e+01	7.1e-10	1.4e-09	-2.7e-09	0.0	0.2
johnson32-2-4.clq	123256	14880	122761	13	4	-3.10000e+01	-3.10000e+01	2.6e-09	2.6e-09	1.2e-08	0.3	1.2
johnson8-2-4.clq	406	168	379	13	5	-7.00000e+00	-7.00000e+00	5.2e-09	2.6e-09	-3.9e-09	0.0	0.1
johnson8-4-4.clq	2485	560	2416	18	5	-5.00000e+00	-5.00000e+00	6.1e-09	1.4e-09	1.4e-09	0.0	0.1
keller4.clq	14706	5100	14536	76	21	-1.50000e+01	-1.50000e+01	3.7e-08	4.2e-07	-1.4e-07	0.5	2.0
keller5.clq	301476	74710	300701	144	44	-3.10000e+01	-3.09999e+01	2.1e-08	4.8e-07	1.8e-06	11.5	42.3
keller6.clq	5649841	1026582	5646481	373	128	-6.30000e+01	-6.29998e+01	1.2e-08	3.6e-07	1.9e-06	1335.3	4849.8
p-hat1000-1.clq	500500	377247	499501	13757	10149	-8.94782e+01	-8.94779e+01	2.7e-08	7.3e-07	1.5e-06	2380.9	8509.9
p-hat1000-2.clq	500500	254701	499501	53723		-6.08624e+01	-6.08623e+01	3.3e-08	4.4e-07	6.4e-07	10302.3	37681.2
p-hat1000-3.clq	500500	127754	499501	746		-1.81366e+01	-1.81366e+01	3.2e-08	3.5e-07	-9.8e-08	129.9	472.3
p-hat1500-1.clq	1125750	839327	1124251	3728	1797	-1.09265e+02	-1.09265e+02	2.5e-08	6.4e-07	7.1e-07	1788.9	6371.0
p-hat1500-2.clq	1125750	555290	1124251	52849		-7.33087e+01	-7.33086e+01	2.6e-08	5.0e-07	8.2e-07	27181.2	99882.0
p-hat1500-3.clq	1125750	277006	1124251	952		-2.14138e+01	-2.14138e+01	2.5e-08	3.8e-07	-4.8e-08	452.6	1657.3
p-hat300-1.clq	45150	33917	44851	2307		-4.40680e+01	-4.40679e+01	6.2e-08	4.6e-07	7.9e-07	44.7	175.6
p-hat300-2.clq	45150	22922	44851	14396		-2.89008e+01	-2.89008e+01	5.3e-08	4.7e-07	4.7e-07	264.9	1044.4
p-hat300-3.clq	45150	11460	44851	1297		-1.04261e+01	-1.04261e+01	5.5e-08	3.9e-07	-3.8e-08	22.8	90.2
p-hat500-1.clq	125250	93181	124751	4365	2479	-5.72741e+01	-5.72740e+01	4.5e-08	6.1e-07	1.2e-06	169.6	655.8
p-hat500-2.clq	125250	61804	124751	53048		-3.90749e+01	-3.90748e+01	4.3e-08	4.8e-07	9.3e-07	2190.6	8566.1
p-hat500-3.clq	125250	30950	124751	674		-1.28264e+01	-1.28265e+01	4.0e-08	3.7e-07	-8.6e-08	25.7	100.3
p-hat700-1.clq	245350	183651	244651	4630	259	-7.34972e+01	-7.34970e+01	4.4e-08	5.0e-07	1.5e-06	403.4	1519.2
p-hat700-2.clq	245350	122922	244651	70731		-5.20063e+01	-5.20062e+01	4.0e-08	4.3e-07	7.0e-07	5840.0	22022.9
p-hat700-3.clq	245350	61640	244651	1788		-1.52407e+01	-1.52407e+01	3.8e-08	3.9e-07	-3.1e-08	136.6	513.2
san1000.clq	500500	249000	499501	233	10	-6.70000e+01	-6.70000e+01	2.5e-08	4.7e-07	-2.1e-08	41.3	149.9
san200-0.7-1.clq	20100	5970	19901	1598	773	-9.01999e+00	-9.01999e+00	5.3e-08	3.2e-07	4.3e-07	14.7	58.4
san200-0.7-2.clq	20100	5970	19901	125		-1.20000e+01	-1.20000e+01	5.4e-08	3.5e-07	1.4e-08	1.2	4.7
san200-0.9-1.clq	20100	1990	19901	13593		-4.01912e+00	-4.01912e+00	6.9e-08	1.9e-07	3.0e-07	123.8	495.1
san200-0.9-2.clq	20100	1990	19901	469	206	-4.30141e+00	-4.30140e+00	5.2e-08	2.0e-07	9.6e-07	4.1	16.5
san200-0.9-3.clq	20100	1990	19901	3283		-5.00000e+00	-5.00000e+00	7.0e-08	2.3e-07	-2.3e-08	29.6	118.3
san400-0.5-1.clq	80200	39900	79801	312	134	-3.42059e+01	-3.42059e+01	4.9e-08	5.0e-07	-1.8e-07	8.4	32.1
san400-0.7-1.clq	80200	23940	79801	229	120	-1.26148e+01	-1.26148e+01	3.8e-08	3.5e-07	5.1e-09	5.5	21.3
san400-0.7-2.clq	80200	23940	79801	351	80	-1.50000e+01	-1.50000e+01	4.6e-08	3.9e-07	1.3e-06	8.9	34.5
san400-0.7-3.clq	80200	23940	79801	107	25	-1.90000e+01	-1.90000e+01	4.2e-08	3.6e-07	-1.7e-07	2.7	10.3
san400-0.9-1.clq	80200	7980	79801	404	111	-5.20011e+00	-5.20011e+00	4.7e-08	2.1e-07	2.1e-07	9.8	37.6
sanr200-0.7.clq	20100	6032	19901	203	91	-8.82923e+00	-8.82923e+00	6.2e-08	4.3e-07	8.1e-08	1.7	6.9
sanr200-0.9.clq	20100	2037	19901	294	23	-4.47656e+00	-4.47656e+00	5.5e-08	2.4e-07	-3.5e-08	2.8	11.1
sanr400-0.5.clq	80200	39816	79801	89	29	-2.01700e+01	-2.01700e+01	4.6e-08	4.2e-07	-9.5e-10	2.4	9.1
sanr400-0.7.clq	80200	23931	79801	89	34	-1.19950e+01	-1.19950e+01	5.2e-08	3.6e-07	-3.9e-08	2.2	8.5

Table 23: Algorithm-Performance NCQNM(15): Table-LovaszSchrijver-dimacs2nd-t=0.95-maxr=15-maxtime=Inf-maxit=100000-tol=1e-06

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf erstellt worden ist.

Felix Lieder, Juni 2018, Düsseldorf