

Diss. ETH No. 20409

First-Order Methods in Large-Scale Semidefinite Optimization

A dissertation submitted to
ETH ZURICH

for the degree of
DOCTOR OF SCIENCES

presented by
MICHAEL KLAUS BÜRGISSEER
Dipl. Math., University of Zurich
born on September 24, 1982
citizen of Willisau (LU)

accepted on the recommendation of
Prof. Dr. Hans-Jakob Lüthi (examiner)
Prof. Dr. Yurii Nesterov (co-examiner)
Dr. Michel Baes (co-examiner)

2012

Abstract

SEMIDEFINITE Optimization has attracted the attention of many researchers over the last twenty years. It has nowadays a huge variety of applications in such different fields as Control, Structural Design, Statistics, or in the relaxation of hard combinatorial problems. In this thesis, we focus on the practical tractability of large-scale semidefinite optimization problems. From a theoretical point of view, these problems can be solved by polynomial-time Interior-Point methods approximately. The complexity estimate of Interior-Point methods grows logarithmically in the inverse of the solution accuracy, but with the order 3.5 in both the matrix size and the number of constraints. The later property prohibits the resolution of large-scale problems in practice.

In this thesis, we present new approaches based on advanced First-Order methods such as Smoothing Techniques and Mirror-Prox algorithms for solving structured large-scale semidefinite optimization problems up to a moderate accuracy. These methods require a very specific problem format. However, generic semidefinite optimization problems do not comply with these requirements. In a preliminary step, we recast slightly structured semidefinite optimization problems in an alternative form to which these methods are applicable, namely as matrix saddle-point problems. The final methods have a complexity result that depends linearly in both the number of constraints and the inverse of the target accuracy.

Smoothing Techniques constitute a two-stage procedure: we derive a smooth approximation of the objective function at first and apply an optimal First-Order method to the adapted problem afterwards. We present a refined version of this optimal First-Order method in this thesis. The worst-case complexity result for this modified scheme is of the same order as for the original method. However, numerical results show that this alternative scheme needs much less iterations than its original counterpart to find an approximate solution in practice. Using this refined version of the optimal First-Order method in Smoothing Techniques, we are able to solve randomly generated matrix

saddle-point problems involving a hundred matrices of size $12'800 \times 12'800$ up to an absolute accuracy of 0.0012 in about four hours.

Smoothing Techniques and Mirror-Prox methods require the computation of one or two matrix exponentials at every iteration when applied to the matrix saddle-point problems obtained from the above transformation step. Using standard techniques, the efficiency estimate for the exponentiation of a symmetric matrix grows cubically in the size of the matrix. Clearly, this operation limits the class of problems that can be solved by Smoothing Techniques and Mirror-Prox methods in practice. We present a randomized Mirror-Prox method where we replace the exact matrix exponential by a stochastic approximation. This randomized method outperforms all its competitors with respect to the theoretical complexity estimate on a significant class of large-scale matrix saddle-point problems. Furthermore, we show numerical results where the randomized method needs only about 58% of the CPU time of the deterministic counterpart for solving approximately randomly generated matrix saddle-point problems with a hundred matrices of size 800×800 .

As a side result of this thesis, we show that the Hedge algorithm – a method that is heavily used in Theoretical Computer Science – can be interpreted as a Dual Averaging scheme. The embedding of the Hedge algorithm in the framework of Dual Averaging schemes allows us to derive three new versions of this algorithm. The efficiency guarantees of these modified Hedge algorithms are at least as good as, sometimes even better than, the complexity estimates of the original method. We present numerical experiments where the refined methods significantly outperform their vanilla counterpart.

Zusammenfassung

DIE Semidefinite Optimierung hat in den letzten 20 Jahren das Interesse unzähliger ForscherInnen auf sich gezogen und weist heutzutage eine Vielzahl von Anwendungen in den verschiedensten Gebieten wie Regelungstechnik, Baukonstruktion oder Statistik auf. Ausserdem werden schwierige kombinatorische Probleme häufig durch semidefinite Hilfsprobleme approximiert. Der Fokus dieser Arbeit liegt auf dem möglichst raschen Lösen von grossen semidefiniten Optimierungsproblemen. Theoretisch können diese Probleme approximativ mit der Inneren-Punkte Methode gelöst werden, da dieser Algorithmus eine polynomielle Laufzeit hat. Die schlimmst mögliche theoretische Laufzeit dieser Methode wächst logarithmisch mit dem Inversen der Fehlertoleranz und mit der Potenz 3.5 bezüglich der Grösse der Entscheidungsmatrix und der Anzahl der Nebenbedingungen. In der Praxis ist tatsächlich eine relativ schnelle, obwohl polynomielle, Zunahme der Rechenzeit bezüglich der Matrixgrösse und der Anzahl der Nebenbedingungen beobachtbar. Dadurch ist die Innere-Punkte Methode ungeeignet für die numerische Lösung grosser semidefiniter Optimierungsprobleme.

In dieser Arbeit werden neue Methoden vorgestellt, die leicht strukturierte, grosse semidefinite Optimierungsprobleme bis auf einen moderaten Approximationsfehler lösen können. Diese Methoden basieren auf fortgeschrittenen Subgradienten Algorithmen wie Smoothing Techniques oder Mirror-Prox Methoden, die aber nur auf Probleme mit einer ganz speziellen Struktur angewendet werden können. Jedoch weicht die Form der semidefiniten Optimierungsprobleme, die in dieser Dissertation betrachtet werden, von der vorausgesetzten Struktur ab. In einem ersten Schritt werden die Ausgangsprobleme daher in eine passende Form umgewandelt und als Sattelpunktprobleme formuliert, auf welche die fortgeschrittenen Subgradienten Methoden angewendet werden können. Die theoretischen Laufzeiten der so erhaltenen Methoden wachsen linear in der Anzahl der Nebenbedingungen und im Inversen der Fehlergenauigkeit.

Smoothing Techniques bestehen aus zwei Schritten. Zuerst wird eine ausreichend differenzierbare Approximation der Zielfunktion hergeleitet. Danach wendet man auf das neue Problem eine optimale Subgradienten Methode an. In dieser Dissertation wird eine verfeinerte Version dieser optimalen Subgradienten Methode hergeleitet. Die theoretische Laufzeit des adaptierten Algorithmus ist im ungünstigen Fall von der gleichen Grössenordnung wie die Laufzeit der ursprünglichen Methode. In der Praxis ist allerdings eine signifikante Beschleunigung beobachtbar, da der angepasste Algorithmus bedeutend weniger Iterationen als die ursprünglichen Methode braucht. Durch die Integration dieser abgeänderten Methode in Smoothing Techniques sind zufällig erzeugte Instanzen von Sattelpunktproblemen mit Matrizen der Grösse $12'800 \times 12'800$ bis auf einen absoluten Fehler von 0.0012 in ungefähr vier Stunden lösbar.

Wenn die fortgeschrittenen Subgradienten Methoden auf die Sattelpunktprobleme, die durch die Transformation der ursprünglichen semidefiniten Optimierungsprobleme entstehen, angewendet werden, muss pro Iteration mindestens ein Matrixexponential berechnet werden. Die Rechenzeit für eine solche Operation wächst kubisch mit der Grösse der symmetrischen Matrix, vorausgesetzt, dass das Exponential über die Diagonalisierung der symmetrischen Matrix berechnet wird. Diese Operation wird in der Praxis offensichtlich zum kritischen Faktor, falls Smoothing Techniques oder Mirror-Prox Methoden zum Lösen von grossen Instanzen der betrachteten Sattelpunktprobleme eingesetzt werden sollen. In dieser Dissertation wird eine randomisierte Mirror-Prox Methode präsentiert, bei der das Matrixexponential durch eine stochastisch erzeugte Approximation ersetzt wird. Dieser randomisierte Algorithmus unterbietet alle existierenden Methoden bezüglich der theoretischen Laufzeit auf einer signifikanten Subklasse von grossen Sattelpunktproblemen mit Matrizen. Diese theoretischen Resultate werden durch Beobachtungen in der Praxis belegt: Die randomisierte Methode benötigt etwa 58% der Rechenzeit des deterministischen Pendant für das approximative Lösen von zufällig erzeugten Instanzen mit Matrizen der Grösse 800×800 .

Als Nebenresultat dieser Dissertation wird der Hedge Algorithmus – eine Methode, die in der Theoretischen Informatik weit verbreitet ist – im Kontext von Dual Averaging Methoden eingebettet und diskutiert. Basierend auf dieser Interpretation werden drei neue Versionen des ursprünglichen Hedge Algorithmus hergeleitet. Die theoretischen Komplexitätsresultate der neu entwickelten Algorithmen sind gleich gut oder gar besser als die Effizienzgarantien der bisherigen Methode. Es werden numerische Resultate präsentiert, bei denen diese neuen Methoden zu deutlich besseren Resultaten führen als der ursprüngliche Hedge Algorithmus.