

Computational complexity certification of gradient methods for real-time model predictive control

Doctoral Thesis

Author(s):

Richter, Stefan

Publication date:

2012

Permanent link:

<https://doi.org/10.3929/ethz-a-007587480>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Diss. ETH No. 20718

Computational Complexity Certification of Gradient Methods for Real-Time Model Predictive Control

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by

Stefan Richter

Dipl.-Ing., Graz University of Technology
born 13.01.1981
citizen of Austria

accepted on the recommendation of

Prof. Dr. Manfred Morari, examiner
Prof. Dr. Yurii Nesterov, co-examiner

2012

Abstract

The main objective of this thesis is to derive practically relevant certificates for the computational complexity of real-time model predictive control (MPC). MPC is a well-established method for the control of systems that are subject to constraints in the inputs and/or states. Since MPC requires the solution to an optimization problem at the sampling rate of the control loop, it is considered a computationally intensive control method. Even for the class of *convex* optimization problems, which are solvable by state-of-the-art optimization methods, the complexity of obtaining a solution is either unknown *a priori* or does not correspond to the practically observed complexity at all. However, a certified operation of the optimization routine is relevant in view of MPC applications for systems with fast dynamics that require sampling times in the milli- to microseconds range, e.g., grid power converters.

The complexity of an optimization method is best expressed by an upper bound on the required number of floating point operations that in turn stems from an upper bound on the iteration count. Knowing a *meaningful* upper bound *a priori* would enable control engineers to use optimization-based control in safety-critical and risk-of-loss applications, guide the selection of control hardware before implementation and reduce extensive and costly testing procedures of the optimization routine. Even if certification is not an issue in an application, the understanding of the entities that define (expressive) iteration bounds can help to modify the problem at hand in a way such that faster practical convergence can be achieved, saving computational resources.

This thesis investigates certification aspects for the most common class of constrained linear-quadratic MPC problems. After confirming that existing state-of-the-art solution methods for MPC do not allow for meaningful complexity certificates, Nesterov's fast gradient method is proposed as an alternative. The fast gradient method is an accelerated variant of the classic gradient method and provides a significant speed-up at the same iteration cost. Moreover, its simple algorithmic scheme and its applicability to a wide class of optimization problems make it a viable optimization method for low-resource processors such as found in embedded systems. Although the method has existed since the early 80s, it was practically unknown in the control community. In this sense, this work has also contributed to making it known to a wider audience and other control labs have started projects on the use of the fast gradient method in the meantime.

In order to account for that and since a similar overview of the method is missing in the literature, the first part of the thesis provides a self-contained treatment of the fast gradient method and its latest generalization, the (accelerated) proximal gradient method. From this it becomes clear that no other gradient-based method can provide better certificates for MPC and also, that the fast gradient method's certificates are indeed practically relevant for the considered MPC problems in this thesis. In addition, two new stopping criteria for gradient-based methods are derived that allow to detect sufficient optimality even before the certified iteration bound is reached.

In the second part, we start with certificates for input-only constrained MPC, which – after changing the problem formulation – can be conveniently solved by the fast gradient method in the primal domain. The validity of the obtained certificates crucially depends on the initialization of the fast gradient method. Therefore, a generic initialization scheme is derived and specialized variants for cold- and warm-starting deduced. Computation of the 'best' bound on the iteration count turns out to be computationally expensive as it requires the solution to a bilevel optimization problem. However, an easy-to-compute, while still expressive bound can be derived for cold-starting. In order to understand for which MPC problem setups low bounds and thus short solution times can be expected, their dependence on the problem data is analyzed, and it is found that the certificates do not necessarily get worse when the prediction horizon in MPC is increased, which at first seems counterintuitive. The application of the certification scheme to several real-world MPC problems, such as a power converter, proves the practical relevance of the obtained bounds and shows that the fast gradient method is the method of choice for most of these problems.

In the input- and state-constrained MPC case, the fast gradient method is shown to be applicable in the dual domain only. This implies that the guaranteed convergence rate changes from linear in the primal to the slower sublinear rate in the dual domain. Reduced convergence speed means that the entities defining the upper iteration bound need to be known with the least amount of conservatism in order to result in a meaningful bound. Motivated by this, the derivation of the *smallest* values of these entities is considered. It turns out that the first entity, which is the smallest Lipschitz constant of the dual gradient, can be computed exactly whereas the other entity, the worst case minimal distance between an initial dual iterate and a Lagrange multiplier, can only be upper bounded with much conservatism. The conservatism causes the bounds for a real-world ball on plate system to be too large to be practical. For this reason, a theoretical framework for the investigation of a less conservative bound is introduced and it is found that under specific assumptions improved upper bounds can be found.

Finally, optimal preconditioning for better certificates and increased computational speed is studied for both the primal and the dual domain.

Zusammenfassung

Ziel dieser Dissertation ist es, praxisrelevante Zertifikate für die Komplexität von modellprädiktiven Reglern (engl.: model predictive control, MPC) zu erhalten. MPC ist ein etabliertes Verfahren zur Regelung von Systemen mit Beschränkungen auf Eingangs- und/oder Zustandsgrößen. Da es die Lösung eines Optimierungsproblems zu jedem Abtastzeitpunkt des Regelkreises erfordert, zählt es zu den rechenintensiven Regelungsverfahren. Selbst für die Klasse der *konvexen* Optimierungsprobleme, die durch moderne iterative Lösungsmethoden als effizient lösbar gelten, ist die Komplexität des Lösungsprozesses entweder *a priori* gänzlich unbekannt oder wird von der im Vorhinein bestimmten Komplexität massiv überschätzt. In Anbetracht vermehrt eingesetzter MPC Regler für hochdynamische Systeme, zum Beispiel Stromwandler im Versorgungsnetz, mit Abtastzeiten im Milli- bis Mikrosekundenbereich, stellt sich jedoch zunehmend die Frage nach einer aussagekräftig zertifizierbaren Lösungsmethode.

Die Komplexität einer Lösungsmethode wird am geeignetsten durch eine Schranke auf die erforderliche Anzahl von Fließkommaoperationen angegeben. Diese wird ihrerseits wiederum von einer oberen Schranke auf die Anzahl der Iterationen der Lösungsmethode abgeleitet. Das *a priori* Wissen um so eine *aussagekräftige* obere Schranke würde eine Vielzahl von Vorteilen mit sich bringen: Einsatz von optimierungsbasierter Regelung in sicherheitskritischen Anwendungen sowie Anwendungen mit hohem Verlustrisiko; Auswahlhilfe für Regelungshardware vor der konkreten Implementierung; Reduktion von umfangreichen und kostspieligen Prüfverfahren der Lösungsmethode selbst. Auch wenn der Zertifizierungsaspekt in einer Anwendung nicht im Mittelpunkt steht, kann das Verständnis aussagekräftiger oberer Iterationsschranken dazu ausgenutzt werden, das ursprüngliche Problem dahingehend abzuändern, dass schnellere Konvergenz und damit Einsparungen bei den Rechenressourcen erzielt werden können.

Diese Dissertation untersucht Zertifizierungsaspekte für die praktisch am häufigsten auftretenden linear-quadratischen MPC Probleme mit Beschränkungen. Nachdem festgestellt wird, dass sich die aktuell im Einsatz befindlichen Lösungsmethoden nicht für eine aussagekräftige Zertifizierung eignen, wird Nesterov's schnelle Gradientenmethode als Alternative vorgeschlagen. Die schnelle Gradientenmethode ist eine Variante der klassischen Gradientenmethode mit einer deutlich erhöhten Konvergenzgeschwindigkeit bei gleichbleibendem

Aufwand pro Iteration. Ihr bestechend einfacher Algorithmus macht die schnelle Gradientenmethode zu einer praktikablen Lösungsmethode für eine grosse Klasse von Optimierungsproblemen, speziell auf Prozessoren mit beschränkter Rechenleistung, wie sie etwa in eingebetteten Systemen zu finden sind. Obwohl die Methode bereits Anfang der 80er Jahre entwickelt wurde, war sie in der Regelungstechnik bis dato weitgehend unbekannt. In diesem Sinne hat diese Arbeit auch dazu beigetragen, sie einem breiteren Kreis von Anwendern zugänglich zu machen. In der Zwischenzeit haben bereits einige andere Forschungsgruppen auf dem Gebiet der Regelungstechnik die schnelle Gradientenmethode als ein interessantes Forschungsfeld entdeckt.

Um dem Rechnung zu tragen und weil eine ähnliche Übersicht in der Literatur noch fehlt, stellt der erste Teil der Dissertation eine in sich geschlossene Abhandlung über die schnelle Gradientenmethode und ihrer neuesten, verallgemeinerten Variante, der (beschleunigten) Proximalgradientenmethode bereit. Hieraus wird ersichtlich, dass keine andere gradientenbasierte Methode bessere Zertifikate für MPC bieten kann, aber auch, dass die schnelle Gradientenmethode praxisrelevante Zertifikate für die in dieser Arbeit betrachteten MPC Probleme zulässt. Zusätzlich werden noch zwei neue Abbruchkriterien für gradientenbasierte Methoden abgeleitet, die es erlauben, hinreichende Optimalität festzustellen, noch bevor die zertifizierten Iterationsschranken erreicht sind.

Der zweite Teil der Arbeit widmet sich zunächst der Berechnung von Zertifikaten für MPC Probleme mit Beschränkungen ausschliesslich auf den Eingangsgrössen. Nach einer Transformation können diese Probleme unmittelbar durch die schnelle Gradientenmethode im primalen Raum gelöst werden. Die Gültigkeit der erhaltenen Zertifikate hängt entscheidend von der Initialisierung der Methode ab, weshalb zuerst eine generische Strategie zur Initialisierung eingeführt wird, um dann in weiterer Folge zwei spezielle Varianten für Kalt- und Warmstart davon abzuleiten. Dabei stellt sich heraus, dass die Berechnung der 'besten' Iterationsschranke die aufwändige Lösung eines Bilevel-Optimierungsproblems voraussetzt. Es wird allerdings gezeigt, dass man eine einfach zu berechnende, dennoch ausdrucksstarke obere Schranke für den Fall einer Kaltstart-Initialisierung erhalten kann. Um einen Anhaltspunkt dafür zu bekommen, welche MPC Konfigurationen kleine obere Iterationsschranken und somit kurze Lösungszeiten bedingen, wird deren Abhängigkeit von den ursprünglichen Problem Daten analysiert. Im Zuge dessen wird unter anderem festgestellt, dass – entgegen der Intuition – die Schranken nicht unbedingt grösser werden, wenn das Prädiktionsfenster des MPC Reglers vergrössert wird. Die Anwendbarkeit des entwickelten Zertifizierungsschemas wird anhand zahlreicher praktischer MPC Beispiele, etwa für einen Stromwandler, untermauert und es wird aufgezeigt, dass für die meisten dieser Beispiele, die schnelle Gradientenmethode die bevorzugte Lösungsmethode darstellt.

Für den allgemeinen Fall von MPC mit beschränkten Eingangs- und Zustandsgrößen wird zunächst festgehalten, dass die schnelle Gradientenmethode nur im dualen Raum anwendbar ist. Dies impliziert jedoch, dass sich die garantierte Konvergenzrate von linearer im primalen Raum zu sublinearer Rate im dualen Raum verschlechtert. Die reduzierte Konvergenzgeschwindigkeit bedeutet, dass jene Größen, die die Iterationschranken für die schnelle Gradientenmethode bestimmen, mit kleinstmöglichem Konservatismus ermittelt werden müssen, um aussagekräftige Schranken zu gewährleisten. Motiviert durch diese Beobachtung beschäftigt sich die Dissertation in diesem Teil mit der Bestimmung der *kleinsten* Werte dieser zentralen Größen. Wie sich herausstellt, kann die eine der beiden Größen, die kleinste Lipschitz-Konstante des dualen Gradienten, genau berechnet werden, während die andere, der grösste minimale Abstand zwischen einem Initialisierungspunkt und einem Lagrange-Multiplikator, nur konservativ nach oben abgeschätzt werden kann. Dieser Konservatismus bewirkt, dass die Iterationsschranken für ein reales Kugel-auf-Platte-System zu gross sind, um praktisch relevant zu sein. Um dem Abhilfe zu verschaffen, wird im Rahmen einer theoretischen Untersuchung eine weniger konservative Lösung skizziert, die unter geeigneten Annahmen in der Tat die Bestimmung besserer oberer Iterationsschranken erlaubt.

Schlussendlich werden noch sowohl für den primalen als auch für den dualen Raum optimale Vorkonditionierungen ermittelt, mit dem Ziel, bessere Zertifikate und kürzere Lösungszeiten zu erhalten.