Vorausschauend planen, gezielt handeln – über die Regelung und Steuerung technischer Prozesse

Lars Grüne* Sebastian Sager†

Frank Allgöwer‡ Hans Georg Bock§ Moritz Diehl¶

15. August 2008

1 Executive Summary

Seit Beginn der industriellen Revolution nimmt die Steuerungs- und Regelungstechnik eine Schlüsselstellung in vielen technischen Bereichen ein. James Watts Fliehkraftregler für Dampfmaschinen ist eines der frühen Beispiele eines extrem erfolgreichen Reglerkonzepts, von dem Ende der 1860er Jahre geschätzte 75000 Exemplare allein in England im Einsatz waren [2, p. 24]. Etwa um diese Zeit begannen Ingenieure, motiviert durch die immer höhere Komplexität der zu regelnden Maschinen, sich systematisch mit theoretischen Grundlagen der Regelung zu beschäftigen. Dies führte unausweichlich zu der Einsicht, dass das dynamische Verhalten der geregelten Systeme nur mit Hilfe der Mathematik verstanden und weiterentwickelt werden konnte, oder wie Werner von Siemens, ein weiterer technischer Pionier in diesem Bereich es formulierte: „Ohne Mathematik tappt man doch immer im Dunkeln."

*Mathematisches Institut, Universität Bayreuth, 95440 Bayreuth, E-Mail lars.gruene@uni-bayreuth.de

†Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, 69120 Universität Heidelberg, E-Mail sebastian.sager@iwr.uni-heidel.berg.de

‡Institut für Systemtheorie und Regelungstechnik, Universität Stuttgart, 70550 Stuttgart, E-Mail allgower@ist.uni-stuttgart.de

§Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, 69120 Universität Heidelberg, E-Mail bock@iwr.uni-heidel.berg.de

¶Optimization and Engineering Center, KU Leuven, E-Mail Moritz.Diehl@esat.kuleuven.be

In der Steuerungs- und Regelungstechnik stellt sich daher – heute ebenso wie vor mehr als hundert Jahren – nicht die Frage, *ob* man Mathematik benutzen muss, sondern vielmehr, *welche* Mathematik man sinnvollerweise benutzen sollte. Tatsächlich gibt es von der Algebra über die Geometrie und die Theorie der Dynamischen Systeme bis hin zur Optimierung und Numerik kaum ein mathematisches Gebiet, das nicht bedeutende Beiträge zur Regelungs- und Steuerungstechnik und ihren mathematischen Nachbardisziplinen System- und Kontrolltheorie geleistet hätte.

In diesem Artikel wollen wir zum einen anhand zweier Beispiele darstellen, welche Beiträge die Mathematik in der langen Geschichte der Steuerungs- und Regelungstechnik geliefert hat und dabei aufzeigen, welche Faktoren grundlegend für den Erfolg dieser Methoden waren. Zum anderen wollen wir – neben einem Überblick über den Status Quo in der Steuerungs- und Regelungstechnik – eine moderne Regelungsmethode, die *Modellprädiktive Regelung* (kurz: MPC, vom englischen „model predictive control“), genauer vorstellen, die dahinterstehenden mathematischen Konzepte aus der Systemtheorie, der Numerik und der Optimierung kurz erläutern und darauf aufbauend einige der zukünftigen Herausforderungen an die Mathematik skizzieren.

In ihrer linearen Variante ist MPC speziell in der Verfahrenstechnik bereits heute ein Standardwerkzeug in industriellen Prozessen, in ihrer nichtlinearen Variante (NMPC), auf die wir hier besonders eingehen werden, gilt sie allgemein als eine der vielversprechendsten modernen Methoden zur Regelung komplexer technischer Prozesse, die sich durch die rasanten Fortschritte sowohl auf theoretischer als auch auf algorithmischer Seite zunehmend neue Anwendungsgebiete erschließt. Zudem ist sie ein Paradebeispiel für eine Methode, die nur durch die interdisziplinäre Zusammenarbeit auf den Gebieten Regelungstechnik, mathematische Systemtheorie, Numerik und Optimierung ihren heutigen Entwicklungsstand erreichen konnte und auch nur durch diese Zusammenarbeit weiter entwickelt werden kann. Diese Interdisziplinarität, die bereits jetzt erfreulicherweise an einer Reihe von Universitäten durch Exzellenzcluster, Forschungszentren und Graduiertenschulen in Forschung und Lehre gefördert wird, erscheint uns als ein wesentlicher Faktor, um die vielfältig vorhandenen Chancen des Produktionfaktors Mathematik im Bereich der Steuerungs- und Regelungstechnik auch in Zukunft nutzen zu können.

2 Eine lange Erfolgsgeschichte

Mathematische Methoden spielen in der Regelungstechnik seit Beginn dieses Gebietes eine wesentliche Rolle. Zwei prominente Beispiele, die wir in diesem Abschnitt kurz erläutern, sollen dies belegen. Anhand dieser Beispiele – das Stabilitätskriterium von Hurwitz und das Pontrjaginsche Maximumprinzip – werden wir prototypisch aufzeigen, welche „Erfolgsfaktoren“ diese beiden mathematischen Entwicklungen so erfolgreich gemacht haben.

2.1 Das Stabilitätskriterium von Hurwitz

Das Stabilitätskriterium von Hurwitz wurde 1893–1894 von dem am Züricher Polytechnikum (heute ETH Zürich) als Professor tätigen Mathematiker Adolf Hurwitz entwickelt [30]. Anlass für die Entwicklung war aber weniger Hurwitz' mathematische Neugier als eine konkrete Anfrage seines Züricher Kollegen Aurel Stodola, eines Maschinenbauingenieurs, der sich mit der Entwicklung von Regelungen hydraulischer Turbinen beschäftigte. Stodolas Problem lässt sich am einfachsten anhand eines alltäglichen Problems veranschaulichen. Stellt man sich eine Heizungsregelung vor, bei der die Raumtemperatur auf einen gewünschten Wert eingestellt werden soll, so ist eine naheliegende Strategie, das Ventil für den Warmwasserdurchfluss im Heizkörper abhängig von der aktuell gemessenen Temperatur einzustellen: Ist die Temperatur geringer als die gewünschte, wird man das Ventil aufdrehen, ist sie höher, wird man es zudrehen. Man benötigt eigentlich keine Mathematik, um es einleuchtend zu finden, dass dieses Vorgehen tatsächlich die gewünschte Temperatur liefern wird.

Allerdings gilt das nur im Idealfall. Komplizierter wird die Sache, wenn der Einfluss der Stellgröße (im Beispiel das Durchflussventil) auf den Wert (also die Temperatur) weniger direkt wird, z.B. durch Verzögerungen im System. Wer schon einmal unter einer Dusche gestanden ist, deren Warmwasserzufuhr nur verzögert auf das Öffnen oder Schließen der Ventile reagiert, kennt diesen Effekt: Statt der gewünschten Temperatur erhält man ständig zwischen „zu heiß“ und „zu kalt“ schwankende Werte – ein klassisches Beispiel für ein instabiles Verhalten eines Regelkreises – und nur mit viel Mühe und Fingerspitzengefühl gelingt es manchmal, am Ende doch die richtige Temperatur einzustellen. Auch wenn dies nur eine Analogie ist, ähnlich sind die Effekte bei den mechanischen Turbinensystemen, die Stodola betrachtete, dort hervorgerufen durch die Vielzahl mechanischer Kopplungen, die zu berücksichtigen sind. Auch hier ist es kompliziert, eine Regelung zu entwerfen, die das System stabil auf einem vorgegebenen Wert hält, es also

stabilisiert.

Stodola kannte und verwendete nun aber einige zu der Zeit knapp 20 Jahre alte mathematische Resultate [62], die es erlaubten, aus einem Modell des Regelkreises eine Bedingung für die Stabilität abzulesen. Diese Bedingung lautete, dass ein Polynom, also eine mathematische Funktion der Form

$$P(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

nur Nullstellen (also komplexe Zahlen x mit $P(x) = 0$) mit negativem Realteil besitzt. Wenn solch ein Polynom nur wenige Summanden besitzt, so ist diese Bedingung leicht nachzuprüfen. Solche einfachen Polynome gehören allerdings auch zu sehr einfachen Regelungssystemen, bei denen diese Theorie gar nicht nötig ist, um die Stabilität sicher zu stellen, da ein erfahrener Ingenieur dies den Systemen sowieso „ansieht“. Wirklich relevant für die Praxis sind Polynome mit vielen Termen und das Hauptproblem ist dann, dass man die Nullstellen – aus prinzipiellen algebraischen Gründen, die seit dem frühen 19. Jahrhundert durch die Arbeiten von Évariste Galois bekannt sind – nicht einfach ausrechnen kann.

Hurwitz beschäftigte sich also auf Anregung von Stodola eine Weile mit diesem Problem und fand schließlich ein Kriterium, mit dem sich ermitteln ließ, ob die Nullstellen negative Realteile besitzen, ohne dass man diese tatsächlich ausrechnen muss. Mathematisch ausgedrückt bestand seine Lösung darin, die Koeffizienten $a_0, a_1, \ldots, a_n$ des Polynoms in n verschiedene Matrizen unterschiedlicher Größe zusammenzufassen. Dann genügt es, lediglich die Determinanten dieser Matrizen zu berechnen, um die Frage nach dem Vorzeichen der Realteile der Eigenwerte zu beantworten.

Stodola war von dieser Lösung begeistert. In einem Brief an Hurwitz schrieb er: „Ich bin Ihnen außerordentlich verpflichtet für die geniale Lösung des Wurzelproblems, das mich so sehr plagte.“ [5]. Aber auch Hurwitz erwähnte in einer Fußnote zu seiner zugehörigen Veröffentlichung in den Mathematischen Annalen nicht ganz ohne Stolz die erfolgreichen praktischen Anwendungen seines Kriteriums: „Herr Stodola benutzt mein Resultat in seiner Abhandlung [...], deren Ergebnisse bei der Turbinenanlage des Badeortes Davos mit glänzendem Erfolge Anwendung gefunden haben.“ Bemerkenswert ist in diesem Zusammenhang noch die Tatsache, dass der englische Mathematiker Edward John Routh bereits knapp 20 Jahre vor Hurwitz eine ähnliche Lösung entwickelte, weswegen Hurwitz' Kriterium heutzutage meist als Methode von Routh-Hurwitz bezeichnet wird. Den Weg in die Anwendungen hingegen hat Hurwitz' Resultat früher gefunden, weil Rouths Ergebnis Jahrzehnte lang nur einem kleinen akademischen Zirkel aber kaum einem in der Praxis arbeitenden Ingenieur bekannt war [2, p. 81f.].

2.2 Das Pontrjaginsche Maximumprinzip

Das Pontrjaginsche Maximumprinzip hat trotz seines auf den russischen Mathematiker Lew S. Pontrjagin zurückgehenden Namens mehrere Urheber. Zunächst einmal hat Pontrjagin das im Jahre 1956 veröffentlichte Prinzip nicht alleine, sondern mit wesentlichen Beiträgen der oft nicht genannten Mathematiker V.G. Boltjansky und R.L. Gamkrelidze veröffentlicht und bewiesen [8] (siehe auch [51]). Zum anderen, vgl. [49, p. 55f.], findet sich das Prinzip bereits in Veröffentlichungen des amerikanischen Mathematikers M.R. Hestenes aus dem Jahre 1950 und – sogar in verallgemeinerter, allerdings nicht vollständig bewiesener Form – des ebenfalls amerikanischen Mathematikers R.P. Isaacs aus dem Jahre 1954/55. Wichtige grundlegende Ideen zum Maximumprinzip findet man sogar noch früher in Carathéodorys Buch von 1935, wie in [50] ausgeführt wird.

Beim Maximumprinzip geht es um die Frage, wie man die Bewegung eines Systems – etwa die Flugbahn einer Rakete oder die Bewegung des Arms eines Industrieroboters – optimal vorausplant. „Optimal“ ist dabei immer auf ein vorher zu definierendes Kriterium mit zugehörigen Nebenbedingungen bezogen, also zum Beispiel: wie kommt eine Rakete mit minimaler Energie (Kriterium) in einer vorgegebenen Zeit (Nebenbedingung) von der Erde in eine spezielle Umlaufbahn, oder wie kann man den Arm eines Industrieroboters schnellstmöglich (Kriterium) bei konstantem Energieaufwand (Nebenbedingung) von einer bestimmten Position in eine andere steuern? Das Beispiel der Rakete ist dabei nicht zufällig gewählt, weil die Raumfahrt ebenso wie die militärische Raketentechnik im beginnenden kalten Krieg zwischen der UdSSR und den USA einer der Hauptgründe für die Entwicklung dieser Art von Mathematik war. Die für die industrielle Entwicklung sicherlich ungleich bedeutendere Anwendung auf Industrieroboter kann daher – ähnlich wie die berühmte Teflon-Pfanne – durchaus als ein Abfallprodukt der Weltraumforschung angesehen werden.

Optimierungsprobleme ähnlicher Art wurden bereits etwa 250 Jahre vor Pontrjagin und den anderen Entwicklern des Maximumprinzips mit Hilfe der sogenannten Variationsrechnung gelöst. Die vollständige Darstellung dieser Methode überschreitet sicherlich den Umfang dieses Beitrags, zumindest eine kurze Beschreibung der prinzipiellen Funktionsweise soll aber hier gegeben werden (für eine detailliertere Darstellung siehe z.B. [49]). Ausgangspunkt für den Lösungsansatz ist zunächst ein Modell, mit dessen Hilfe eine Bewegung durch mathematische Gleichungen beschrieben werden kann. Dies geschieht üblicherweise mit Hilfe eines Systems von Differentialgleichungen. Zur Lösung des Problems muss nun für jeden Zeitpunkt die Position auf der

optimalen Bahn berechnet werden. Das Problem dabei ist nun, dass die Wahl der Werte zu verschiedenen Zeitpunkten voneinander abhängt: Beschleunigt man z.B. am Anfang sehr stark, so muss man am Ende stärker bremsen und umgekehrt. Die Variationsrechnung löste diese komplexen Abhängigkeiten durch das Aufstellen eines erweiterten Differentialgleichungssystems (die Euler-Lagrange-Gleichungen), dessen erster Teil „normal“ vorwärts, dessen zweiter Teil aber rückwärts gelöst wird – ein sogenanntes Randwertproblem. Für jeden Zeitpunkt kann aus der Lösung dieses Systems dann die optimale Bahn ermittelt werden.

Das Hauptproblem der Anwendung der Variationsrechnung auf praktische Probleme besteht nun darin, dass dieses direkt die optimalen Bahnen berechnet, nicht aber die eigentlich relevanten optimalen Steuerwerte, die an die Raketentriebwerke oder die Roboterarmmotoren übermittelt werden müssen – und weil diese Werte in der Rechnung nicht auftauchen, ist es auch unmöglich, Wertebereiche vorzugeben, in denen diese Steuerwerte schon aus physikalischen Gründen liegen müssen. So kann die Variationsrechnung zwar durchaus optimale Flugbahnen ermitteln, es ist aber unmöglich auszuschließen, dass diese eine Schubkraft der Triebwerke benötigen, die weit jenseits der physikalischen Möglichkeiten liegt. Dies ist ganz offensichtlich eine Unzulänglichkeit, die nicht nur in der Raumfahrt, sondern auch in fast jeder denkbaren industriellen Anwendung ernsthafte praktische Probleme bereitet.

Hier bietet nun das Maximumprinzip Abhilfe: Da hier – bei konzeptionell durchaus ähnlicher Verwendung von Randwertproblemen – der Fokus nicht auf der Berechnung optimaler Bahnen, sondern auf der direkten Berechnung der zugehörigen optimalen Steuerstrategien liegt, ist es nun leicht möglich, physikalisch oder auch ökonomisch notwendige Beschränkungen direkt in die Rechnung einzubeziehen. Dabei liefert das Maximumprinzip ein durchaus auch intuitiv verständliches Kriterium, da die optimalen Steuerwerte in jedem Zeitpunkt nichts anderes als die Lösung eines neuen „kleinen“ Optimierungsproblems sind, dessen Parameter sich gerade aus dem gegebenen Randwertproblem ergeben. Da sich solche Probleme auch numerisch am Rechner gut lösen lassen, wird das Prinzip nicht nur für die theoretische Analyse, sondern auch als Basis für Algorithmen verwendet.

2.3 Fazit

Die in diesem Abschnitt beschriebenen mathematischen Konzepte zur Analyse und Lösung von Steuerungs- und Regelungsprobleme sind nur zwei von vielen möglichen Beispielen. Sie besitzen aber eine Reihe von Eigenschaf-

ten, die in mehr oder weniger starker Ausprägung typisch für erfolgreiche mathematische Konzepte sind:

- Sie sind anwendbar, ohne dass man die zu ihrer Herleitung nötigen tieferliegenden mathematischen Theorien verstehen muss,
- sie haben für reale Anwendungen einen deutlichen Fortschritt gebracht und damit neue industrielle Entwicklungen erst ermöglicht,
- sie sind konstruktiv in dem Sinne, dass sie sich leicht als Algorithmus formulieren lassen und daher mit der Verfügbarkeit digitaler Rechner auch leicht implementiert werden konnten.

Bei all den offensichtlichen Vorteilen, die diese mathematischen Konzepte den industriellen Anwendungen gebacht haben, wären deren Entwicklungen aber trotzdem unmöglich gewesen, wenn bei den in den Steuerungs- und Regelungstechnik tätigen Ingenieuren nicht von vornherein ein Bewusstsein für die Notwendigkeit mathematischer Methoden – und damit einhergehend eine solide mathematische Grundbildung – vorhanden gewesen wäre. Ohne Stodolas Kenntnisse über die Bedeutung der Nullstellen wäre Hurwitz vermutlich nie auf die Idee gekommen, sein Kriterium zu entwickeln und ohne die Verfügbarkeit geeigneter Differentialgleichungsmodelle wäre die Entwicklung des Maximumprinzips bereits im Ansatz unmöglich gewesen.

3 Status Quo und aktuelle Entwicklungen: Das Beispiel „Modellprädiktive Regelung“

3.1 Eine kurze Einführung in die Steuerungs- und Regelungstechnik

Das Ziel von Steuerungen und Regelungen ist die Beeinflussung dynamischer Systeme, so dass deren Verhalten gewünschte Eigenschaften besitzt. Die dynamischen Systeme werden dabei durch sogenannte *Stellgrößen* beeinflusst. Beim dynamischen System „Auto“ sind dies z.B. der Lenkradeinschlag oder die Stellung des Gaspedals. Durch geeignete Veränderung der Stellgröße über der Zeit kann dann die Systemdynamik in der gewünschten Weise beeinflusst werden. Das interessierende Verhalten des zu beeinflussenden Systems wird oft in der sogenannten *Regelgröße* zusammen gefasst. Bild 3.1 zeigt schematisch eine solche *Regelstrecke*.

Man unterscheidet nun zwischen Steuerung und Regelung. Wie in Bild 3.2 zu sehen ist wird bei einer *Steuerung* das dynamische Verhalten durch eine

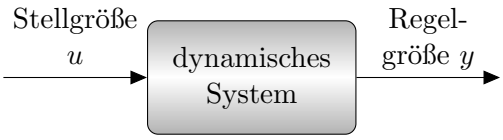

Abbildung 3.1: Schematische Darstellung einer Regelstrecke mit Stellgröße u und Regelgröße y.

offene Wirkungskette beeinflusst. Der Sollverlauf für die Regelgröße y wird durch die *Sollgröße* w vorgegeben, die den Eingang für die Steuerung bildet. Die Steuerung generiert aus diesem Signal den Verlauf der Stellgröße, die dann das dynamische System so beeinflusst, dass die Regelgröße dem Verlauf der Sollgröße möglichst exakt folgt.

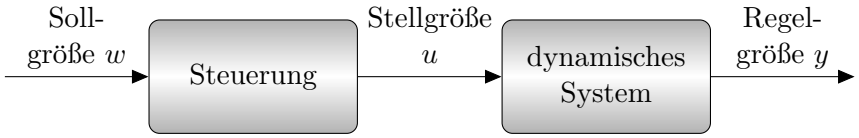

Abbildung 3.2: Wirkprinzip der Steuerung.

Steuerungen sind in der industriellen Praxis sehr wichtig und erfüllen ihre Aufgabe zumeist sehr gut. Als Beispiele seien hier Robotersteuerungen oder Steuerungen für Werkzeugmaschinen und Fertigungseinrichtungen genannt. Grundlegende Voraussetzung für die korrekte Funktion einer Steuerung ist aber, dass das Verhalten des dynamischen Systems mit Hilfe des vorhandenen Modells hinreichend genau vorhergesagt werden kann; ist dies nicht der Fall, so kann das tatsächliche Verhalten weit von der zur Berechnung der Stellgröße u verwendeten theoretischen Vorhersage abweichen.

Dies kann z.B. dann passieren, wenn große externe Störungen auf das dynamische System wirken, wie z.B. starker Seitenwind bei Automobilen, oder wenn das Verhalten des dynamischen Systems nicht exakt bekannt ist, weil das mathematische Modell nicht alle Aspekte des realen Systems abbildet – in der Regelungstechnik spricht man in diesem Fall von *unsicheren Systemen*. Die Auswirkungen solcher Unsicherheiten sind immer dann besonders gravierend, wenn das Systemverhalten *instabil* ist, wenn also wie beim Beispiel der Dusche in Abschnitt 2.1 kleine Änderungen in der Stellgröße (Ventil) zu großen Änderungen in der Regelgröße (Wassertemperatur) führen. In diesen Fällen müssen *Regelungen* eingesetzt werden, um die Stabilität des Systems – also die Tatsache, dass die Regelgröße stabil in der

Nähe des Sollwertes bleibt – zu erreichen.

Eine Regelung besitzt im Gegensatz zu einer Steuerung eine Rückführung (engl. feedback), wie in Bild 3.3 gezeigt. Die Regelgröße wird gemessen, zurückgeführt und mit ihrem gewünschten Verlauf, also der Sollgröße, verglichen. Die Differenz zwischen Soll- und Istwert, auch Regelfehler genannt, wird nun dem Regler als Eingang zur Verfügung gestellt. Durch den geschlossenen Wirkungskreis kann der Einfluss von externen Störungen und Unsicherheiten durch die Regelung explizit erkannt und korrigiert werden, bevor diese zu großen Abweichungen vom Sollwert führen. Daher funktionieren Regelungen auch bei Vorhandensein von Störungen, Unsicherheiten und bei instabilen Regelstrecken.

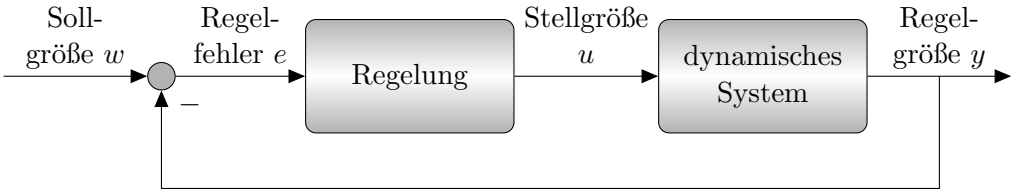

Abbildung 3.3: Wirkprinzip einer Regelung mit Rückführung und Soll-/Istwert-Vergleich.

Natürlich führt die Rückführungsstruktur nicht automatisch zur Stabilität des geschlossenen Wirkungskreises: Entscheidend ist, nach welcher Vorschrift die Stellgröße aus dem Regelfehler berechnet wird. Wie in Abschnitt 2 bereits beispielhaft erläutert, kann das dynamische Verhalten eines geregelten oder gesteuerten mit den mathematischen Methoden der System- und Regelungstheorie untersucht werden. Die wichtigsten Analyseverfahren untersuchen dabei die Stabilität und die sogenannte *Robustheit der Stabilität* des geregelten oder gesteuerten Systems. Bei der Untersuchung der Robustheit der Stabilität wird die Frage beantwortet, ob die Stabilität erhalten bleibt, wenn sich das wirkliche dynamische System von den im Modell gemachten Annahmen unterscheidet. Diese Eigenschaft ist für den Einsatz der Methoden in der praktischen Anwendung von zentraler Bedeutung.

Von ganz entscheidender Relevanz ist die Mathematik beim *Entwurf* der Steuerungen und vor allem der Regelungen, d.h. bei der Herleitung der Vorschrift, mit der die Stellgröße aus der Sollgröße oder dem Regelfehler berechnet wird. Moderne Regler werden heute *modellbasiert* entworfen. Das heißt es wird angenommen, dass ein mathematisches Modell des zu regelnden Systems, zumeist in Form von Differentialgleichungen, vorliegt. Weiterhin werden die Ziele der Regelung ebenfalls mathematisch formuliert. Das ist zum

Einen die bereits mehrfach genannte Stabilität, aber auch die sogenannte *Regelgüte*, die zumeist in Form von Optimierungskriterien formuliert wird. Unterschiedliche Klassen von Reglerentwurfsverfahren unterscheiden sich, erstens, durch unterschiedliche Annahmen an die dem Entwurf zugrundeliegenden Modelle der Regelstrecken und, zweitens, durch unterschiedliche Zielvorgaben. In Bezug auf die Modellannahmen ist die wichtigste Unterscheidung diejenige in lineare bzw. nichtlineare Systeme. Lineare Systeme zeichnen sich – etwas vereinfacht gesagt – dadurch aus, dass der Zusammenhang zwischen Stellgröße u und Regelgröße y proportional ist, dass also zum Beispiel doppelt so große Stellgrößen genau doppelt so große Regelgrößen nach sich ziehen (mathematisch sagt man, sie erfüllen das Superpositionsprinzip). Sie werden durch lineare Differentialgleichungen, zumeist in der sogenannten Zustandsraumdarstellung, beschrieben

$$\dot{x} = Ax + Bu \tag{3.1a}$$

$$y = Cx + Du. \tag{3.1b}$$

Der Entwurf von Reglern für lineare Systeme ist sehr gut verstanden und es stehen eine große Anzahl von äußerst leistungsfähigen linearen Reglerentwurfsverfahren zur Verfügung. An dieser Stelle seien beispielhaft die LQ-Verfahren erwähnt (LQR, LQG, LQG/LTR, etc.) [36, 42, 23, 47], bei denen für lineare Systeme der Form (3.1) ein Regler entworfen wird, so dass der geschlossene Kreis optimales Verhalten in Bezug auf die Minimierung eines quadratischen Integralkriteriums

$$J = \int_0^\infty x^T(\tau)Qx(\tau) + u^T(\tau)Ru(\tau)\,\mathrm{d}\tau$$

besitzt.

Bei den LQ-Verfahren werden keine externen Störungen, wie zum Beispiel der Seitenwind beim Auto, berücksichtigt. Beim moderneren H_∞-Reglerentwurf [24, 64] werden die Regler dagegen genau so entworfen, dass sie externe Störgrößen optimal unterdrücken, d.h. dass deren negative Auswirkungen auf den Regelkreis minimiert werden. Diese Minimierung kann über unterschiedliche mathematische Optimierungskriterien erfolgen und neben den H_∞-Verfahren können hier beispielhaft auch die L_1-Methoden [14] genannt werden.

Auch wenn kaum ein in der industriellen Praxis auftretendes System tatsächlich linear ist, können lineare Modelle trotzdem oft eingesetzt werden. Der Grund dafür liegt darin, dass sich die allermeisten Systeme „fast“ wie lineare Systeme verhalten, wenn man nur einen kleinen Wertebereich

für die Regelgröße betrachtet: Will man die Raumtemperatur von 20° auf 22° Celsius erhöhen, so muss man das Heizungsventil in etwa zweimal so weit öffnen, als wenn man die Temperatur nur auf 21° erhöhen möchte – der Zusammenhang zwischen Ventilöffnung und Temperaturzunahme ist linear. In größeren Temperaturbereichen gilt das aber nicht mehr, schon allein deswegen, weil sich das Ventil irgendwann nicht mehr weiter aufdrehen lässt und die Maximaltemperatur erreicht ist.

Wenn das zu regelnde System im relevanten Wertebereich kein lineares Verhalten aufweist, was in der Praxis häufig der Fall ist, müssen nichtlineare Reglerentwurfsmethoden angewandt werden. Das dem Entwurf zugrunde liegende dynamische System wird hier durch nichtlineare Differentialgleichungen der Form

$$\begin{aligned} \dot{x} &= f(x,u) \\ y &= h(x,u). \end{aligned}$$

beschrieben. Auf dem Gebiet der nichtlinearen Regelung wurden in den letzten ca. 20 Jahren sehr große Fortschritte gemacht und viele aktuelle Forschungsprojekte befassen sich mit diesem Thema. Beispielhaft sollen hier die differentialgeometrischen Verfahren der exakten Linearisierung [33], die flachheitsbasierte Regelung [22, 58], die passivitätsbasierte Regelung [9, 55], oder die Backstepping-Methoden [37, 41] genannt sein. Diese Methoden liefern im einfachsten Fall[1] eine statische Funktion

$$u = k(x), \tag{3.3}$$

d.h. eine mathematische Formel $k(x)$, die abhängig vom aktuellen Systemzustand x angibt, wie die Stellgröße u gewählt werden muss.

Im Nichtlinearen steht häufig das Stabilisierungsproblem im Vordergrund. Selbstverständlich ist aber auch hier die Regelgüte von Interesse. Das nichtlineare Äquivalent zu den oben erwähnten LQ-Verfahren ist die optimale Regelung, bei der wieder ein Optimierungskriterium und Nebenbedingungen – wie bereits in Abschnitt 2.2 beschrieben – vorgegeben werden. Im Falle der nichtlinearen optimalen Regelung ist die Berechnung einer expliziten Formel (3.3) für u in der Praxis aber selbst mit Hochleistungscomputern oft unmöglich. Einfacher ist es hingegen, nichtlineare optimale Steuerungen (also Stellgrößen u in Abhängigkeit von der Zeit) mit dem Computer zu berechnen, z.B. mit Hilfe des Pontrjaginschen Maximumprinzips, vgl. Abschnitt 2.2 oder der moderneren direkten Methoden, vgl. Abschnitt 3.5.

[1] Oft ist der Regler nicht durch eine statische Gleichung, sondern selbst wieder durch eine Differentialgleichung gegeben.

Deshalb wurde in den letzten Jahren eine neue Klasse von Verfahren zur optimalen Regelung nichtlinearen Systeme entwickelt, bei denen keine explizite Formel der Form (3.3) berechnet wird, sondern bei denen das Rückführgesetz $k(x)$ online in Echtzeit aus optimalen Steuerungen berechnet wird. Obwohl solch eine online-Rechnung auf den ersten Blick wesentlich schwieriger aussieht, zeigt es sich, dass diese Vorgehensweise deutliche Vorteile hat und zur Lösung von großen nichtlinearen Regelproblemen in der Praxis geeignet ist. Der prominenteste Vertreter dieser neuen Klasse von Regelungsverfahren ist die modellprädiktive Regelung, die in den folgenden Abschnitten genauer betrachtet und diskutiert werden soll.

3.2 Prinzip der modellprädiktiven Regelung

Wenn man einen Regler entwirft, der z.B. die Temperatur eines Hauses auf einem möglichst konstanten Wert halten soll, so besteht die einfachste Idee darin, die aktuelle Temperatur zu messen und darauffolgend die Warmwasserdurchflussmenge passend zu erhöhen oder zu verringern – wie in Abschnitt 2.1 erläutert, kann man bei geschickter Wahl der Regelparameter dadurch ein stabiles geregeltes System erreichen, das mehr oder weniger gut die durch das Wetter verursachten äußeren Temperaturschwankungen auffängt. Da die Heizung nicht beliebig schnell Wärme nachliefern kann, und auch die Abkühlung nicht beliebig schnell folgen kann, wird die Raumtemperatur dabei immer leicht schwanken.

Was wäre aber, wenn wir den Wetterbericht nutzen würden? Dann könnten wir, wenn eine Kaltfront naht, die Wohnung bereits etwas vorwärmen, oder, wenn warmes Wetter erwartet wird, bereits einige Stunden im Voraus die Heizung abstellen, so dass das Haus sich langsam abkühlen kann. Es ist intuitiv klar, dass die Verwendung dieses „prädiktiven“ Prinzips uns erlaubt, das Regelungsziel „möglichst kleine Abweichungen vom Sollwert“ unter Beachtung unserer begrenzten Heiz- oder Kühlleistung besser einzuhalten.

Wenn wir selbst Auto fahren, ist es sogar absolut unabdingbar, dass wir „vorausschauend“, also „prädiktiv“ fahren, denn würden wir nach schneller gerader Fahrt ohne vorheriges Abbremsen in eine Kurve rasen, flögen wir aus dieser. Dass wir das nicht tun, ist allein der fortgeschrittenen Regelungstechnik in unserem Gehirn zu verdanken, die die Kurvenfahrt bereits antizipiert, lange bevor wir die Fliehkraft in der Kurve am eigenen Leib spüren, siehe Abbildung 3.4.

Die modellprädiktive Regelung (model predictive control, MPC) verwirklicht genau dieses Prinzip für die automatische Regelung technischer Systeme, vgl. z.B. [11, 46]. Wir erläutern dies nun etwas detaillierter am Beispiel

Abbildung 3.4: Vorausschauendes Fahren - so wie es sein soll...

einer typischen Stabilisierungsaufgabe: der Systemzustand soll zu einem Sollwert geregelt und dort gehalten werden, indem der Zustand regelmäßig gemessen wird und gegebenenfalls festgestellte Abweichungen durch geeignete Anpassung der Stellgrößen ausgeglichen werden.

Ausgehend vom aktuellen Zustand x_0 des zu steuernden Systems und den aktuellen Vorhersagen über externe Einfüsse, berechnet ein modellprädiktiver Regler dazu eine (hier der Einfachheit halber stückweise konstante) optimale Steuerung für die nahe Zukunft, den „Voraussagehorizont“ der Länge T. Das dabei verwendete Optimalitätskriterium minimiert typischerweise den Abstand zum Sollwert, damit die berechnete optimale Lösung einen möglichst geringen Abstand von diesem hat. Von dieser Steuerung wird dann nur der erste Abschnitt u_0 an die reale Anlage gegeben, mit dem das System für eine kurze Zeit – die Abtastzeit δ – gesteuert wird. Nach Ablauf dieser kurzen Zeit wird der neue aktuelle Zustand ermittelt, der Vorhersagehorizont wird um δ nach vorne geschoben, und eine neue optimale Steuerung auf diesem Horizont wird berechnet, von dem wieder nur der erste kurze Teil verwirklicht wird. Auf diese Art und Weise werden die online berechneten optimalen Steuerungen zu einer Regelung „zusammengesetzt“. Abbildung 3.5 illustriert einen Schritt dieses Verfahrens.

Die zwei wesentlichen Vorteile der modellprädiktiven Regelungmethode sind die Fähigkeit, vorausschauende Informationen zu berücksichtigen und Optimalitätskriteria und Beschränkungen einzubeziehen. Diese Vortei-

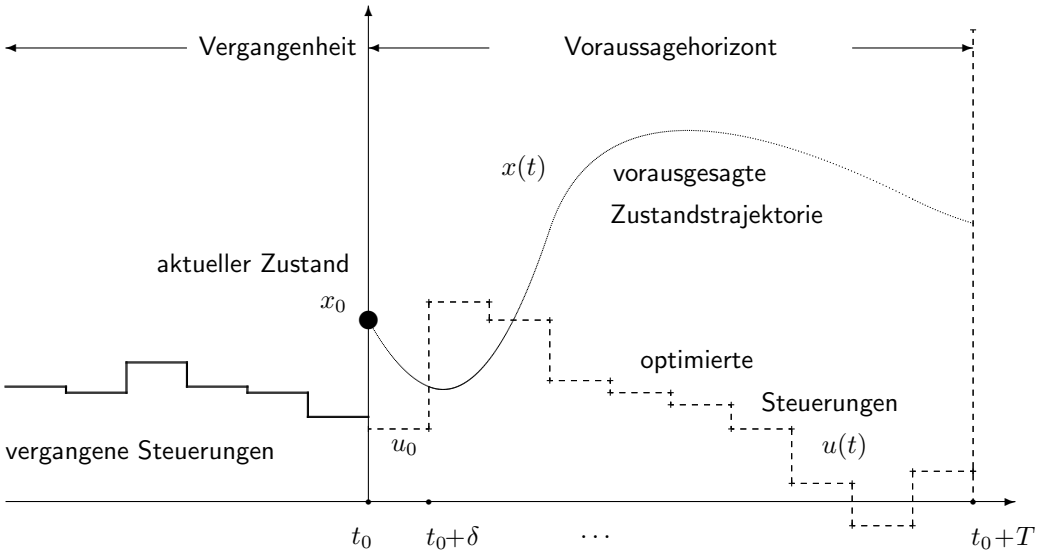

Abbildung 3.5: Schematische Darstellung der modellprädiktiven Regelung.

le kommen z.B. bei energieintensiven oder zeitkritischen Anwendungen zum Tragen, da Ziele wie „mit minimalem Energieaufwand“ oder „innerhalb einer gegebenen Zeit“ direkt in das Optimierungskriterium oder die Nebenbedingungen einbezogen werden können.

Die zwei wichtigsten Fragen sind:

- Wie kann Stabilität des mit MPC geregelten Systems garantiert werden?
- Wie können die Optimalsteuerungsprobleme auf dem Vorhersagehorizont schnell und zuverlässig gelöst werden?

Beide Fragen sind zutiefst mathematischer Natur, die erste fällt in den Bereich der Systemtheorie, die zweite in den Bereich der Numerik. Wir wollen diese Fragen – sowie eine Vorstellung von Anwendungen von MPC – in den folgenden Abschnitten behandeln.

3.3 Stabilität von MPC

Da wir exakte mathematische Definitionen vermeiden wollen, soll „Stabilität“ hier einfach die Tatsache bezeichnen, dass der Regelalgorithmus sein Ziel erfüllt, dass der Systemzustand also in den Sollwert geregelt wird. Intuitiv erscheint es plausibel, dass man durch Minimierung des Abstandes

vom Sollwert eine in diesem Sinne stabile Regelstrategie erhält. Dass das aber durchaus schiefgehen kann, wollen wir an einem kleinen Beispiel veranschaulichen, vgl. Abbildung 3.6. In dieser Abbildung soll die Position des

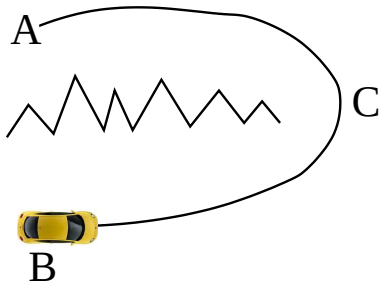

Abbildung 3.6: Beispiel eines Regelungsproblems.

Autos (Systemzustand, zu Beginn im Ort B) in den Ort A (Sollwert) gebracht werden, wobei die Geschwindigkeitsbeschränkungen auf der Straße eingehalten werden sollen. Da ein Gebirge im Weg ist, führt die Straße allerdings nicht auf direktem Weg von A nach B, sondern auf einem Umweg über den Ort C.

Nimmt man als Optimierungskriterium nun den Luftlinien-Abstand des Autos von Punkt A, so muss man diesen Abstand auf dem Weg der Straße von B nach A zunächst vergrößern, um ihn letztendlich verringern zu können. Eine MPC-Strategie, deren Vorhersagehorizont z.B. nur bis zum Punkt C in der Kurve reicht, wird dies nicht erkennen, denn jede Bewegung des Autos im Rahmen dieses Horizonts würde den Abstand ja nur vergrößern, weswegen es optimal ist, einfach im Punkt B stehen zu bleiben. Die geregelte Lösung wird für alle Zeiten im Punkt B verharren und nie im Punkt A ankommen: der Regelalgorithmus ist nicht stabil.

Dieses Problem ist natürlich seit langer Zeit bekannt und es wurden verschiedene Möglichkeiten zur Abhilfe vorgeschlagen. Ein in der Literatur oft betrachteter Ansatz ist es, in der Optimierung auf dem Horizont T nur solche Lösungen zu berücksichtigen, die im gewünschten Sollwert enden. Diese Methode funktioniert theoretisch (was mathematisch rigoros bewiesen wurde, vgl. [38]) und wird in der Praxis auch gelegentlich eingesetzt, besitzt aber den Nachteil, dass man eine weitere Beschränkung zu dem zu lösenden Optimierungsproblem hinzufügt und erfordert zudem i.A. einen sehr großen Horizont T um sicherzustellen, dass es unter den gegebenen Beschränkungen (also z.B. den Geschwindigkeitsbeschränkungen in unserem Straßenbeispiel) überhaupt möglich ist, bis zur Zeit T im Sollwert anzukommen. Beides kann

die Lösung des Optimierungsproblems deutlich erschweren. Ein weiterer Ansatz ist die Verwendung von Korrekturtermen in der Optimierung, die den Effekt von zu kurzen Horizonten T ausgleichen, siehe z.B. [13]. Die Berechnung dieser Terme ist aber i.A. recht aufwendig, weswegen diese Methode in der Praxis eher selten angewendet wird.

Die in der Praxis zumeist verwendete Lösung ist die naheliegendste: Die Horizontlänge T wird einfach so lange erhöht, bis der Algorithmus stabil wird. Interessanterweise wird diese Lösung im Großteil der theoretischen MPC-Literatur *nicht* betrachtet und tatsächlich wurde für nichtlineare MPC-Methoden unter relativ allgemeinen Voraussetzungen erst vor kurzem mathematisch rigoros nachgewiesen, dass dies auch wirklich funktioniert [34, 26]. Nachteil dieser Methode ist, dass im vorhinein i.A. nicht klar ist, wie groß T denn nun gewählt werden muss – und je größer T ist, desto mehr Zeit benötigt die Lösung des Optimierungsproblems, die ja online während der Laufzeit des Systems durchgeführt werden muss. Aktuelle Forschungsansätze [27, 28] versuchen daher, den benötigten Horizont T aus den Systemeigenschaften abzuschätzen und dabei insbesondere zu ermitteln, wie das Optimalitätskriterium gewählt werden muss, damit man mit möglichst kleinem T Stabilität erhält. Über die Praktikabilität dieses Vorgehens für industriell relevante Prozesse kann derzeit nur spekuliert werden. Für unser Straßenbeispiel liefert dieser Ansatz allerdings eine ebenso natürliche wie effiziente Lösung, auf die der eine oder die andere unter den Leserinnen und Lesern vielleicht schon gekommen ist: Misst man den Abstand vom Ort A nicht über die Luftlinie, sondern über die Straßenlänge, so ist es bei beliebig kurzem Horizont T immer besser, sich auf A zuzubewegen, da der Abstand in jedem Fall verringert wird. Wird über diesen Abstand optimiert, so ist das Verfahren daher für beliebige Horizonte T stabil.

3.4 Anwendungsbeispiele

Die Modellprädiktive Regelung (MPC) wurde ursprünglich in der Verfahrenstechnik entwickelt, wo sie z.B. zum Regeln großer und langsamer Destillationskolonnen verwendet wird – diese Systeme sind extrem energieintensiv, und die langsame Zeitskala lässt viel Zeit für die Berechnung der optimalen Steuerungen. Die Verfahrenstechnik ist auch heute noch das Hauptanwendungsgebiet von MPC: Dittmar und Pfeiffer [20] haben 2005 in einer Erhebung 9456 MPC Anwendungen ermittelt (gehen aber von einer deutlich höheren tatsächlichen Zahl aus), von denen mehr als 80% im Bereich der Verfahrenstechnik liegen. Schwerpunktmäßig wird MPC dabei bei der Regelung kontinuierlicher Reaktoren eingesetzt, bei denen der Zustand des

Prozesses (Temperatur, Konzentration der Chemikalien ...) kontinuierlich auf einem Sollwert gehalten werden muss. Seltener wird MPC bei sogenannten Batch-Reaktoren verwendet, bei denen die Reaktion nicht kontinuierlich, sondern in einem vorgebenenen Zeitrahmen stattfindet und bei denen die Regelungsaufgabe darin besteht, einem vorausberechneten Verlauf der Prozesszustände zu folgen.

Wenngleich es Beispiele erfolgreicher Anwendungen nichtlinearer MPC (NMPC) in der Verfahrenstechnik gibt [52], werden hier überwiegend lineare MPC Verfahren eingesetzt, bei denen das zu Grunde liegende Differentialgleichungsmodell linear ist. Dies hat den großen Vorteil, dass die resultierenden Optimierungsprobleme linear sind, wodurch schnellere und zuverlässigere Lösungsalgorithmen zur Verfügung stehen. Zwar hat eine vergleichende Studie an einem industriellen Batch-Reaktor [48] ergeben, dass NMPC-Verfahren durchaus zu einer besseren Regelung führen, die Verbesserungen sind bisher aber noch nicht so ausgeprägt, dass sich die aufwendigere Implementierung des Verfahrens ökonomisch lohnen würde – es bleibt abzuwarten, ob sich dies mit knapper werden Rohstoffen und teurerer Energie in Zukunft ändern wird. Einen weiteren Hinderungsgrund für die Anwendung von NMPC stellt in vielen Fällen das Fehlen geeigneter nichtlinearer Modelle in der Verfahrenstechnik dar [52], ein Bereich, in dem in jüngster Zeit allerdings mit Hilfe geeigneter mathematischer Methoden deutliche Fortschritte erzielt wurden, siehe Abschnitt 4.1.

Durch effizientere Algorithmen und theoretische Fortschritte erschließen sich MPC und insbesondere NMPC-Methoden seit einiger Zeit immer weitere Anwendungsfelder außerhalb der Verfahrenstechnik. Beispiele sind die Regelung von Schleusen in Kanalsystemen, die Regelung von Heizungs- und Klimasystemen in großen Gebäuden, oder die optimale Regelung von saisonalen Wärmespeichern. Letztere nutzen die Wärme im Sommer, um den Grund unter einem Gebäude aufzuheizen und dabei gleichzeitig das Gebäude zu kühlen, und gebrauchen diese Wärme im Winter mit Hilfe einer Wärmepumpe wieder, um Heizkosten zu sparen. Nur die prädiktive Regelung erlaubt es hier, zu entscheiden, wann die Wärmepumpe verwendet werden soll und wann optimalerweise zugeheizt werden soll, basierend auf den saisonalen Witterungsvorhersagen. Zunehmend schnellere Optimierungsalgorithmen erlauben immer schnellere Anwendungen der modellprädiktiven Regelung. So ist heute das punktgenaue Regeln der Fahrt von Zügen möglich und experimentell getestet [25], ebenso wie die Regelung der Einspritz- und Luftmenge bei Automotoren im 50 Millisekundentakt [21]. Auch Roboterarme, die komplexe Manöver in möglichst kurzer Zeit ausführen sollen, und dies in ständig wechselnder Umgebung, sollen demnächst mit MPC geregelt

werden, siehe Abbildung 3.7. Eine der visionärsten Anwendungen der MPC

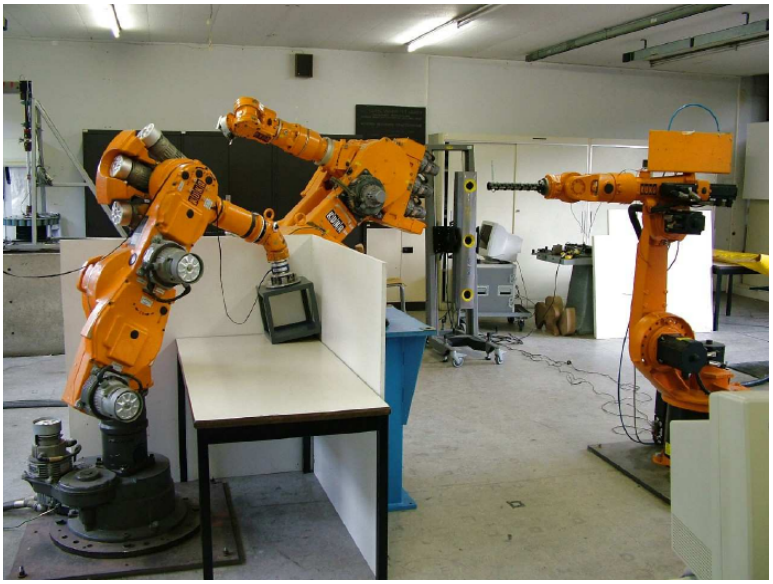

Abbildung 3.7: Roboter an der Universität Leuven, an denen zur Zeit mit MPC im Millisekundentakt experimentiert wird.

ist die automatische Steuerung von Flugdrachen, die in Zukunft Windenergie aus großen Höhen liefern sollen und unter wechselnden Windbedingungen stabil auf Kreisbahnen am Himmel fliegen sollen [12, 32], siehe Abbildung 3.8. Eine weitere zur Zeit untersuchte Anwendung der MPC ist die automatische Steuerung der Insulinzugabe bei Diabetikern.

3.5 Direkte Optimalsteuerungsmethoden

Wie weiter oben angekündigt, wollen wir nun einen kurzen Einblick in den aktuellen Stand der Forschung im Bereich der Methoden zur Lösung des rechenintensivsten Teilschritts der modellprädiktiven Regelung geben, der Lösung der einzelnen Optimalsteuerungsprobleme auf dem betrachteten Zeithorizont. Wir konzentrieren uns dabei auf den für die aktuelle Forschung ungleich relevanteren Fall der nichtlinearen Systeme.

In Abschnitt 2.2 wurde das Maximumprinzip angesprochen. Auf diesem mathematischen Theorem basierende, sogenannte *indirekte* Methoden sind eine Möglichkeit, optimale Lösungen für Steuerungsprobleme zu berechnen.

Abbildung 3.8: Automatisch betriebener Flugdrachen zur Energieerzeugung, [12].

Diese werden in der Tat auch heute noch eingesetzt. Dies gilt insbesondere für die Luft- und Raumfahrt, siehe beispielsweise [10], aber auch für die Analyse von Problemen in der Verfahrenstechnik, siehe [59]. Das Maximumprinzip nimmt auch nach wie vor eine wichtige Rolle in der Analyse von Eigenschaften von Optimalsteuerungsproblemen ein. Für die Konstruktion von Algorithmen, die am Computer umgesetzt werden können, haben diese indirekten Methoden allerdings einige Nachteile:

- der erhebliche analytische Aufwand schon bei kleinen Änderungen des Modells oder von Parametern,
- die komplizierte Behandlung von Pfad- und Steuerungsbeschränkungen, die zu zustandsabhängigen Sprüngen in den differentiellen Variablen führen,
- die häufig numerisch sehr schwierig zu lösenden Randwertprobleme mit extrem kleinen Konvergenzbereich, die nur durch den Einsatz von Homotopien lösbar sind.

Auf Grund dieser Nachteile haben sich für die konkrete Berechnung von Regelungsstrategien in der Praxis zumeist die auch unter dem Schlagwort „first

discretize, then optimize“ bekannt gewordenen *direkten* Methoden durchgesetzt, siehe auch die Diskussion in [4].

Eine weitere wichtige Frage ist der zeitliche Verlauf der berechneten optimalen Steuerfunktionen. Beim Maximumprinzip ist die optimale Steuerung eine fast beliebige Funktion, strukturelle Einschränkungen an ihren Verlauf können dort nicht gemacht werden. Von einem praktischen Gesichtspunkt aus betrachtet macht dies nicht unbedingt Sinn. Man stelle sich beispielsweise einen Temperaturregler vor, dessen Steuergröße sich aus technischen Gründen nicht kontinuierlich, sondern nur in regelmäßigen Abständen ändern lässt. Solch ein Regler kann dann nicht jede beliebige Steuerfunktion exakt nachfahren, sondern diese nur approximieren. Eine solche Approximation ist auch das Grundprinzip der direkten Methoden.

Bei diesen Methoden werden die Steuerfunktionen durch Funktionen, die durch endlich viele Steuergrößen bestimmt sind, approximiert. Im einfachsten Fall sind dies stückweise konstante Funktionen, die also auf einem gegebenen Zeitintervall $[t_i, t_{i+1}]$ den konstanten Wert q_i annehmen. In Abbildung 3.9 wird dies veranschaulicht.

Diese Transformation auf ein mathematisches Optimierungsproblem, das nun nur noch von endlich vielen Freiheitsgraden abhängt (nämlich genau diesen q_i's), erlaubt den Einsatz der weit entwickelten Methoden der nichtlinearen Optimierung (nonlinear programming). Hier konkurrieren im wesentlichen zwei Typen von Algorithmen miteinander: Innere Punkt- und active set basierte Methoden. Während bei letzteren die Menge der aktiven Ungleichungen (diese wird als active set bezeichnet) mitgeführt und von Iteration zu Iteration modifiziert wird, bewegen sich Innere Punkt Methoden durch Hilfe von logarithmischen Barrierefunktionen innerhalb des zulässigen Bereiches auf dem sogenannten zentralen Pfad auf die Lösung zu. Beide Algorithmentypen haben gewisse Vor- und Nachteile, die dazu führen, dass sie für eine bestimmte Klasse von Optimierungsproblemen schneller die optimale Lösung bestimmen. Eine Grundregel ist dabei, dass Innere Punkt Methoden häufig besser bei Problemen mit vielen Ungleichungsbeschränkungen funktionieren (weil sie aus dem Inneren kommen und die vielen Schnittpunkte der Ungleichungsbeschränkungen also gar nicht „wahrnehmen“, die bei anderen Algorithmen der Reihe nach „besucht“ werden müssen), während active set basierte Methoden effizientere Warmstarts in der Lösung verwandter Optimierungsprobleme erlauben. Dieses liegt daran, dass sich die Mengen der aktiven Ungleichungen häufig nur geringfügig unterscheiden, wenn sich beispielsweise durch neue Messwerte ein leicht verändertes Optimierungsproblem ergibt. Ein Start in der schon berechneten Lösung des Ursprungsproblems führt dann häufig schon nach wenigen Iterationen zum Ziel.

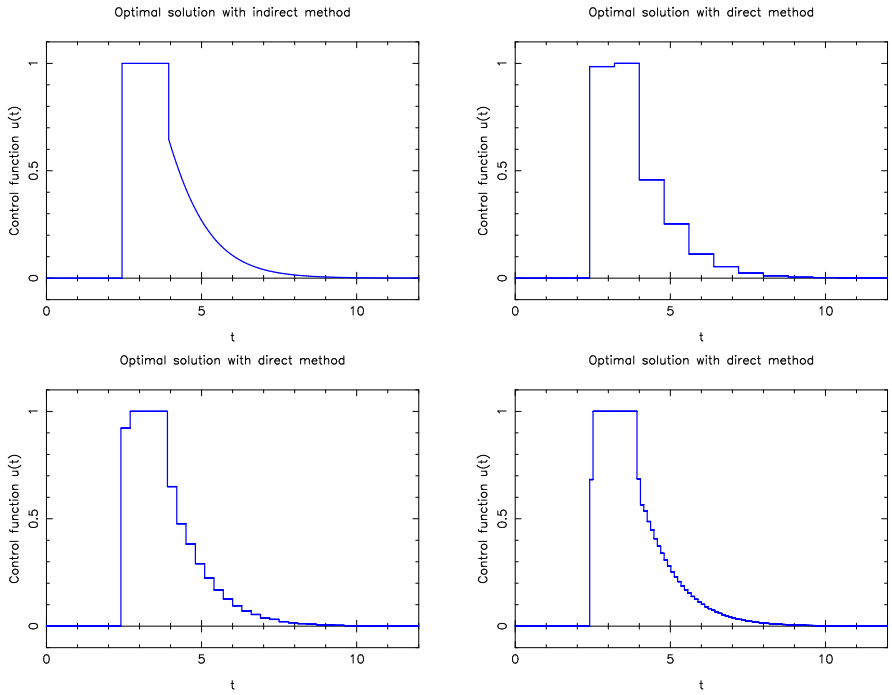

Abbildung 3.9: Links oben: Steuerung mit indirekter Methode berechnet. Die anderen Diagramme zeigen Approximationen mit der direkten Methode auf unterschiedlichen Diskretisierungsgittern.

Bei der Diskretisierung von Optimalsteuerungsproblemen mit direkten Methoden werden gemeinhin drei Algorithmentypen unterschieden, siehe [4]: direktes *single shooting*, direktes *multiple shooting* und direkte *Kollokation*. *Single shooting* wurde in den 70er Jahren entwickelt, siehe beispielsweise [29] und [54]. Die Grundidee besteht darin, die differentiellen Variablen als abhängige Größen der unabhängigen Steuergrößen zu betrachten und in jeder Iteration des Optimierungsalgorithmus ein Anfangswertproblem zu lösen, um die Zielfunktion und Nebenbedingungen sowie die Ableitungen nach den Steuergrößen zu berechnen. Anders ausgedrückt: wir haben einen äußeren Optimierungsalgorithmus, in dem unsere Freiheitsgrade optimiert werden, und einen inneren, in dem wir für festgehaltene Lösungen „schießen“, also durch Integration die Differentialgleichung lösen um den Wert der Zustände zu allen Zeitpunkten zu berechnen. Abbildung 3.10 visualisiert diese Grundidee.

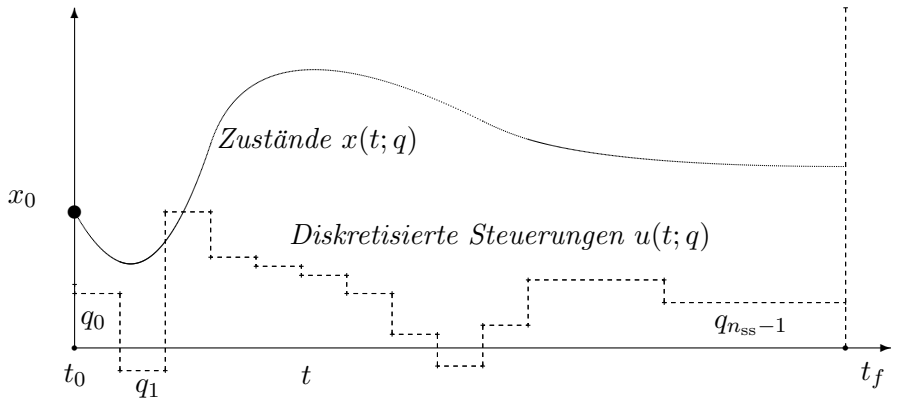

Abbildung 3.10: Schematische Illustration der direkten single shooting Methode. Die Steuerungen sind durch stückweise konstante Funktionen q_i gegeben, die zugehörigen Zustände werden durch Integration ermittelt. Die Intervalllängen müssen nicht notwendigerweise äquidistant sein, sondern können beispielsweise auch auf den letzten Intervallen länger sein, wie angedeutet.

Bei Optimierungsproblemen ist vor allem der Zustandswert am Ende des Zeithorizontes häufig von Interesse. Wenn man an den Wurf eines Balles denkt, so wird schnell klar, wie wichtig die Feinabstimmung aller Steuergrößen ist, möchte man ein Ziel in großer Distanz exakt treffen. Um diese Schwierigkeit über den Horizont zu verteilen, wurde das Konzept des single shooting zum direkten *multiple shooting*, der direkten Mehrzielmethode, erweitert. Diese Methode wurde erstmals 1981 veröffentlicht, [6]. Seither wurde die Methode ständig weiterentwickelt oder neu implementiert, siehe beispielsweise [43].

Wie beim single shooting auch, werden die Kontrollen auf einem Zeitgitter diskretisiert. Um den Ballwurf über eine weite Strecke aber besser koordinieren zu können, werden weitere Mitspieler in Form von zusätzlichen Variablen eingesetzt. Hier muss der Ball nun nur noch zum nächsten Mitspieler „geworfen“ werden, ähnlich wie bei einer Menschenkette zum Löschen eines Feuers. Da Nichtlinearitäten und eventuelle Instabilitäten des Problems auf einem kurzen Zeithorizont nicht so stark zum Tragen kommen wie auf einem langen, ist es nun deutlich einfacher, das anvisierte Ziel zu treffen. Diese Herangehensweise ist ein gutes Beispiel für die Strategie, ein schweres Problem in mehrere kleine aufzuteilen. Dieses Prinzip *divide et impera*, teile

und herrsche, findet sich in vielen mathematischen Algorithmen wieder.

Als neue „Mitspieler“ werden für die Zustände auf einem gegebenen Zeitgitter $t_0 \leq t_1 \leq \cdots \leq t_{n_{\mathrm{ms}}} = t_f$ multiple shooting Variablen $s_i^x \in \mathbb{R}^{n_x}$ mit $0 \leq i < n_{\mathrm{ms}}$ eingeführt, die bei der entkoppelten Integration auf den Zeitintervallen $[t_i, t_{i+1}]$ als Anfangswerte fungieren. Wie in Diagramm 3.11 zu sehen, ist eine derart zusammengesetzte Trajektorie auf dem gesamten Zeithorizont nur dann stetig, wenn die Bedingungen

$$s_{i+1}^x \quad = \quad x(t_{i+1}; s_i^x, q_i, p) \tag{3.4}$$

erfüllt sind. Die linke Seite der Gleichung (3.4) entspricht dem Anfangswert des Zeitintervalls $[t_i, t_{i+1}]$, die rechte Seite dem aufintegrierten Wert des Zustandes zum Zeitpunkt t_{i+1} in Abhängigkeit vom Startwert s_i^x und den Steuergrößen q_i und p. Diese Gleichungen werden dem Optimierungsproblem hinzugefügt und müssen von einer optimalen Lösung erfüllt werden. Aus diesem Grund wird die direkte Mehrzielmethode häufig als *all-at-once* Ansatz bezeichnet, da Simulation und Optimierung zur gleichen Zeit betrieben werden im Gegensatz zum sequentiellen Ansatz des single shooting.

Anschaulich gesprochen muss jeder Mitspieler sicherstellen, dass sein Wurf auch den jeweiligen Nachfolger erreicht, damit der Ball auch wirklich vom Anfangspunkt zum Endpunkt befördert wird. Die Entscheidungen, die jeder Mitspieler zu treffen hat, unsere Steuerungen q_i, müssen natürlich auch im single shooting getroffen werden. Nun kümmert sich aber jeder erstmal vorrangig um seinen eigenen Zeitabschnitt, und darum, wie er den Ball zum Nachbarn bekommt und sich dabei trotzdem optimal verhält. Dieses „sich um seinen Kram kümmern“ ist eine der größten Stärken dieses Ansatzes. Neben dem Zugewinn an Stabilität führt es nämlich dazu, dass sich Strukturen im Optimierungsproblem ergeben, die algorithmisch ausgenutzt werden können. Die Matrix, die sich bei der Analyse der notwendigen Optimalitätsbedingungen ergibt, hat Blockgestalt, wobei die Blöcke alle Variablen umfassen, die zu einem Zeitintervall gehören.

Aufgrund der Parametrisierung des Zustandsraumes sind sich *direkte Kollokation* und die direkte Mehrzielmethode recht ähnlich. Kollokation geht zurück auf [61] und wurde unter anderem in [7], [3] und [56] erweitert. Die Grundidee hier ist, gar keine unabhängigen Integratoren für die Lösung der Differentialgleichung zu nutzen, sondern sämtliche Gleichungen zu diskretisieren und mit in das Optimierungsproblem aufzunehmen.

Kollokation und Mehrzielmethode teilen mehrere Vorteile im Vergleich zum single shooting. So

- ist es möglich, Vorwissen über den Prozess (das beispielsweise in Form

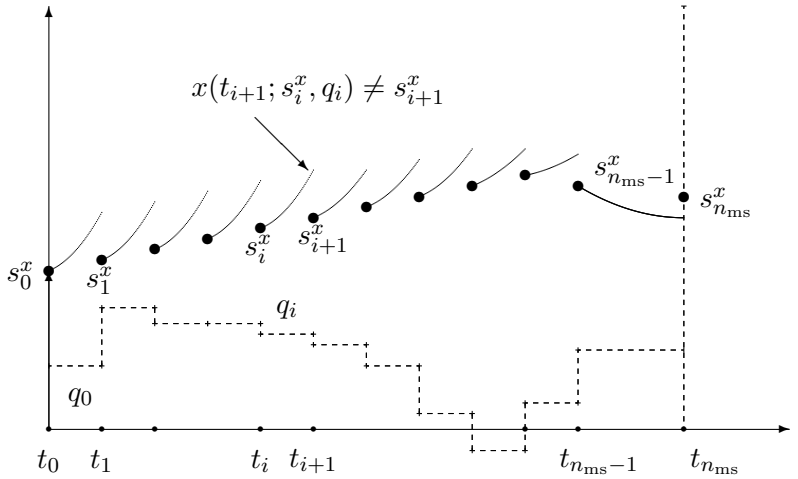

Abbildung 3.11: Schematische Illustration der direkten Mehrzielmethode (multiple shooting). Die Steuerungen sind durch stückweise konstante Funktionen q_i gegeben, die zugehörigen Zustände werden durch stückweise Integration ermittelt. Die Anschlussbedingungen sind in diesem Beispiel nicht erfüllt, die resultierende Trajektorie ist noch unstetig.

von Messdaten vorliegt) zur Initialisierung der Zustandsvariablen zu nutzen,

- ist garantiert, dass die auftretenden Differentialgleichungssysteme auch für nichtlineare Systeme überhaupt lösbar sind, während single shooting für schlecht gewählte Steuerungen in eine Singularität laufen kann,

- können auch instabile Systeme gelöst werden, wenn sie gut gestellt sind, da sich Störungen nicht über den kompletten Zeithorizont ausbreiten, sondern durch die Toleranzen in den Parametrisierungsvariablen gedämpft werden,

- zeigt sich in der Praxis ein wesentlich besseres Konvergenzverhalten im höherdimensionalen Raum.

Die jeweils resultierenden Optimierungsprobleme sind größer als im Fall des single shooting. Dies wird aber durch spezielle strukturausnutzende Algorithmen kompensiert. Im Falle der Kollokation sind dies Lösungsalgorithmen, die die spezielle Bandstruktur der dünnbesetzten Matrizen ausnutzen. Bei der direkten Mehrzielmethode werden Kondensierungsalgorithmen dem

Quadratischen Programm vorgeschaltet, während für die Approximation der Hessematrix sogenannte high-rank updates genutzt werden. Diese tragen der speziellen Blockstruktur durch die Shooting-Intervalle Rechnung und akkumulieren so in jeder Iteration mehr Krümmungsinformation als dies bei normalen Quasi-Newton Methoden der Fall ist.

Zusammenfassend lässt sich sagen, dass sich aus unserer Sicht die direkte Mehrzielmethode und die Kollokation durchgesetzt haben, um nichtlineare Steuerungsprobleme effizient zu lösen. Darüberhinaus können diese Methoden effizient an die spezielle Struktur von NMPC-Anwendungen angepasst werden: Insbesondere die Tatsache, dass hier iterativ eine Serie ähnlicher Optimalsteuerungsprobleme gelöst wird, wurde in den letzten Jahren ausgenutzt, um speziell an NMPC angepasste Varianten dieser Methoden zu entwickeln [44, 18, 57], für die auch mit inakkuraten numerischen Lösungen noch die Stabilität des geregelten Systems nachgewiesen werden kann [17, 19, 63].

4 Herausforderungen

Wir hoffen, geneigte Leserinnen und Leser bis hierher von der bedeutenden Rolle der Mathematik als Schlüsselfaktor für regelungstechnische Produktionsprozesse überzeugt zu haben, und einen – natürlich subjektiven – Einblick in den aktuellen Stand der Forschung und die Vielfalt der Anwendungsgebiete gegeben zu haben. In diesem Kapitel wollen wir uns nun einigen der vor uns liegenden Aufgaben und Herausforderungen widmen. Dazu betrachten wir in Abschnitt 4.1 Aspekte, die sich bei der Modellierung der betrachteten Prozesse ergeben und schauen, inwieweit die Mathematik auch auf dieser Ebene einen gewichtigen Beitrag leisten kann. In Abschnitt 4.2 diskutieren wir die Robustheit von MPC-Regelungen und im abschließenden Abschnitt 4.3 beleuchten wir einige weitere aktuelle Themen.

4.1 Modellierung

Die in den vorgehenden Abschnitten beschriebenen Methoden zur modellgestützten Steuerung von Prozessen haben eine entscheidende Prämisse: *dass sich für den betrachteten Prozess eine adäquate mathematische Beschreibung in Form eines Differentialgleichungssystems finden lässt.*

Die Schlüsselwörter in diesem Satz lassen schon erahnen, wo die Schwierigkeiten stecken.

- „für den betrachteten“ deutet an, dass es um problemspezifische Mo-

dellierungen geht. Von einer Anlage zur nächsten können einzelne Komponenten geringfügig anders sein – aber damit kann der gleiche Prozess ein komplett anderes mathematisches Modell erfordern.

- „adäquat“ impliziert mindestens dreierlei: zum einen, dass das Modell korrekt ist, also die wesentlichen Eigenschaften, die einen im speziellen Zusammenhang interessieren, beinhaltet. Zum zweiten aber auch, dass nicht wesentlich mehr modelliert ist, als unbedingt nötig ist, um Rechenzeiten nicht explodieren zu lassen. Und zum dritten, dass das resultierende Modell überhaupt numerisch geeignet ist für eine Simulation und Optimierung.

- Mit „mathematischer Beschreibung“ sind die unterliegenden Gesetzmäßigkeiten gemeint – ebenso aber auch die Anpassung an das spezifische Problem in Form von konkreten Zahlenwerten, die in das Problem eingehen (häufig als Parameter bezeichnet) und einem Vergleich von Simulationsdaten mit Messwerten standhalten müssen.

- Das sich etwas „finden lässt“, legt eine gewisse Zufälligkeit nahe. Die Frage stellt sich also, wie man solche „Findungsprozesse“ automatisieren kann.

Die mathematische Modellierung beschäftigt sich damit, ein mathematisches Modell für einen realen Vorgang zu erstellen. Dieses basiert auf Wissen aus dem jeweiligen Anwendungsgebiet. So ist spätestens seit Newton bekannt, dass die Beschleunigung die zweite Ableitung der zurückgelegten Strecke ist, welche sich daher bei festem Anfangszustand und gegebener Beschleunigung berechnen lässt. Ein anderes Beispiel sind bekannte Erhaltungssätze, die oftmals physikalischen oder chemischen Modellen zugrunde liegen. Alle solchen Modelle sind natürlich nur Näherungen der Wirklichkeit und vereinfachen an der ein oder anderen Stelle. Von der Berücksichtigung von Reibungskräften bis zu Wechselwirkungen auf Atomebene kann ein solches Modell beliebig verfeinert werden. Eine wichtige Aufgabe des Modellierers ist es, die richtige Mischung zwischen Einfachheit des mathematischen Modells und einer genügend genauen Beschreibung der Wirklichkeit zu finden.

Für viele Probleme der Praxis sind die betrachteten Prozesse leider häufig noch zu komplex, oder nicht ausreichend verstanden. Dies gilt insbesondere für den Bereich der Systembiologie, der sich den Methoden des Wissenschaftlichen Rechnens erst seit einigen Jahren ernsthaft öffnet. Besser ist die Lage in der Robotik, Mechanik, oder Verfahrenstechnik, wo schon seit einigen Jahrzehnten Methoden der Modellierung, Simulation und auch Optimierung Anwendung finden. Die grundlegenden Prinzipien sind hier häufig

sogenannte *first principles*, also grundlegende und gut verstandene Erhaltungssätze der Naturwissenschaften. Leider bleibt trotzdem festzuhalten, dass eine geeignete Modellierung – insbesondere im Bereich der nichtlinearen Modelle – nur durch sehr arbeitsintensive Bemühungen aufbauend auf vorhandenem Expertenwissen gelingt.

Die Verwirklichung der eingangs genannten Prämisse für relevante technische Prozesse bildet daher in der Tat eine große Herausforderung für die Zukunft. Es gibt aber durchaus erste Erfolge auf dem Weg zu einer algorithmischen Modellierung, die wir im Folgenden kurz exemplarisch anreißen wollen – nicht ohne zu betonen, dass auch diese Bewegung weg von dem mühsam und durch trial and error erreichten Modellierungserfolgen im stillen Kämmerlein hin zu standardisierten und verifizierbaren Methoden nicht ohne unterliegende mathematische Methoden denkbar wäre.

Parameterschätzung. Auch wenn die grundlegenden Gesetzmäßigkeiten bekannt sind, so sind doch häufig bestimmte Modellparameter unbekannt. Man denke beispielsweise an Massen von Komponenten in Mehrkörpersystemen, an Aktivierungsenergien, oder an Reaktionsgeschwindigkeiten. Zur Bestimmung dieser Parameter sind Messungen vonnöten, wobei häufig nicht die Parameter direkt gemessen werden können, sondern nur Zustände des Systems, oder Funktionen dieser Zustände. Die Parameterschätzung zielt nun darauf ab, den Abstand zwischen einer Modellantwort und erhobenen Messwerten zu minimieren durch ein *Fitten* der unbekannten Parameter. Die Methode der kleinsten Quadrate (least squares) wurde schon von Carl-Friedrich Gauß und Adrien-Marie Legendre Anfang des neunzehnten Jahrhunderts entwickelt. Sie erfuhr aber in den letzten Jahrzehnten noch bedeutende Weiterentwicklungen, insbesondere was eine effiziente Numerik, Beschränkungen an das Parameterschätzproblem, und robuste Parameterschätzung durch Verwendung von Zielfunktionalen, die Ausreißer zulassen, angeht, siehe beispielsweise [40].

Versuchsplanung. Die Parameterschätzung liefert nicht nur Parameterwerte, für die eine Simulation des Differentialgleichungssystems einen minimalen Abstand zu gemessenen Werten hat. Darüber hinaus erlaubt eine statistische Analyse die Berechnung der Covarianzmatrix, welche ein Maß für die Genauigkeit liefert, mit der die Parameter bestimmt wurden. Vereinfacht gesagt entspricht dies Fehlerbalken, die häufig zu den Messwerten in Diagramme eingezeichnet werden.

Nun ist es so, dass verschiedene experimentelle Versuchsanordnungen zu unterschiedlichen Genauigkeiten führen. Das Ziel der Versuchsplanung ist es, Versuchsanordnungen zu bestimmen, die zu möglichst kleinen Konfidenzge-

bieten (Fehlerbalken) führen. Unter Versuchsanordnung wird hierbei alles verstanden, was ein Experimentator an der Durchführung eines Versuches beeinflussen kann. Dies schließt Steuerungen genauso ein wie die Entscheidung, zu welchen Zeitpunkten (bzw. bei räumlich verteilten Prozessen: an welchen Orten) gemessen werden soll.

Während die Versuchsplanung in der Statistik schon seit langen Jahren untersucht wird, blieb der Transfer in die technischen Anwendungen weitestgehend aus. Dies liegt vor allem darin begründet, dass die Versuchsplanung lange Zeit als theoretisches Konstrukt untersucht wurde, und sich konkrete Rechnungen auf akademische Beispielprobleme beschränkten. Erst in den letzten Jahren wurde an einer effizienten numerischen Umsetzung und an einer Verallgemeinerung auf den nichtlinearen, dynamischen Fall gearbeitet. Die hierbei erzielten Durchbrüche in der Weiterentwicklung der statistischen Analyse erlauben nunmehr den Einsatz selbst innerhalb der industriellen Praxis, wo nichtlineare, dynamische Systeme von bis zu hundert differentiellen Zuständen, oder zunehmend räumlich verteilte Prozesse betrachtet werden.

Insbesondere für Prozesse, deren Dynamik durch Nichtlinearitäten geprägt ist, sind die berechneten optimalen Versuchspläne häufig alles andere als intuitiv, können aber erhebliche Verbesserungen bringen, wie das Beispiel der in Heidelberg am Interdisziplinären Zentrum für Wissenschaftliches Rechnen entwickelten und inzwischen regelmäßig beim Industriepartner BASF in Ludwigshafen eingesetzten Methoden zeigt: Nach Erfahrungswerten in mehr als zwei Dutzend konkreten Projekten zeichnet sich ab, dass gegenüber den traditionellen, von Experten ausgelegten Versuchsanordnungen, Einsparungen von über 80% der Versuche bei gleichbleibender oder gar besserer statistischer Güte möglich sind [1]. In [39] wird von einem optimierten Versuchsplan mit nur 2 Experimenten berichtet, der die zu schätzenden Parameter auf 1% genau bestimmt, während ein zum Vergleich von einem Experten erstellter Versuchsplan mit 15 Experimenten einen Parameter lediglich auf plus minus 30% bestimmt. Gerade in dem hochkompetitiven Umfeld der Verfahrenstechnik tun sich dadurch enorme Einsparpotentiale auf, was Kosten, Umweltbelastung, und time-to-market angeht.

Weitere wichtige Forschungsfelder sind die *Modelldiskriminierung*, die sich mit der automatischen Erstellung von Versuchsplänen beschäftigt, um miteinander konkurrierende Modelle möglichst signifikant voneinander abgrenzen zu können, sowie die *Modellreduktion*, bei der es darum geht, ein schon bestehendes Modell zu vereinfachen, ohne die relevanten Eigenschaften zu verlieren. Ein Ansatz dazu ist die Spektralanalyse der Moden, die langsame Komponenten von den schnellen trennt, welche für die Steuerung

oftmals nur eine untergeordnete Rolle spielen.

4.2 Robustheit von Lösungen, Unsicherheiten

Während (N)MPC ein Konzept zur Regelung von Prozessen ist, das Unsicherheiten implizit Rechnung trägt, so ist insbesondere bei den in Abschnitt 3.5 vorgestellten Methoden zur offline-Optimierung vorausgesetzt, dass der zu optimierende Prozess sich deterministisch so verhält, wie es das mathematische Modell vorhersagt. In der Praxis wird dies nicht der Fall sein, denn es treten verschiedene Arten von Unsicherheiten auf:

- **Modellungenauigkeiten.** Trotz der in Abschnitt 4.1 vorgestellten Bemühungen, ein adäquates Modell für einen betrachteten Prozess zu erlangen, wird es doch immer einen Modellfehler geben, der zu (hoffentlich nur geringfügigen) Abweichungen der Simulationsergebnisse von dem realen Prozess führt.

- **Externe Störungen.** Durch zufallsbedingte Störungen (Rauschen) werden Einflüsse auf das System ausgeübt, die zu dem Zeitpunkt der Optimierung nicht bekannt sein konnten. Beispiele sind Seitenwind beim Auto, Regenschwankungen bei Abwasserkläranlagen, Konsumentenverhalten in der Wirtschaft, oder externe Temperaturschwankungen.

- **Unterschiedliche Szenarien.** Diese Art von Unsicherheit entsteht dadurch, dass bestimmte Parameter des untersuchten Prozesses a priori offen sind. Beispielsweise sollen Turbinenschaufeln oder Flugzeuge gute aerodynamische Eigenschaften für eine ganze Bandbreite von Anströmwinkeln haben. Oftmals unterscheidet man hier noch weiter zwischen einer diskreten Anzahl von Szenarien, einer Wahrscheinlichkeitsverteilung in einem kontinuierlichen Bereich, und einem worst-case Szenario.

Für alle drei Typen von Unsicherheit stellt sich sowohl dem Anwender, wie dem zuarbeitenden Mathematiker die Frage, wie man mit dieser Herausforderung umgeht.

Berechnete optimale Lösungen haben oftmals eine Eigenschaft, die ihren Einsatz in der Praxis unattraktiv macht: sie sind sehr sensitiv gegenüber Unsicherheiten. Dies liegt daran, dass die nichtlinearen Effekte so weit es geht ausgereizt werden. Auch Beschränkungen werden an vielen Stellen aktiv, das heißt maximal ausgereizt. Wenn man zeitoptimal von A nach B fahren

möchte, so wird man an einigen Stellen genau die maximal erlaubte Geschwindigkeit fahren. Wenn nun das Modell ungenau ist, oder aber externe Störungen (man denke an Gegen- bzw. Rückenwind) auftreten, so werden diese Beschränkungen unter Umständen verletzt. Bei besagter Geschwindigkeitsrichtlinie mag das akzeptabel sein, aber bei sogenannten *runaway*-Prozessen, bei denen ab einem gewissen Punkt nicht mehr rückgesteuert werden kann, ist es dies nicht. Wer wollte verantworten, dass in einem Kernkraftwerk die maximale Energieeffizienz erzielt wird, wenn man dafür ständig kurz vor einer unumkehrbaren Kettenreaktion steht?

Für einen praktischen Einsatz hätte man daher viel lieber eine Lösung, die *fast* genau so gut ist wie die optimale, aber weniger störungssensitiv. Ansätze, die in diese Richtung gehen, arbeiten mit a) Feedback, so wie es auch MPC durch das Einarbeiten von Messwerten in einem Echtzeit-Kontext tut, oder b) mit Sicherheitsmargen, die auf die Beschränkungen addiert werden, oder c) mit gewichteten Summen, multikriterieller Optimierung oder spieltheoretischen Ansätzen (hier gibt es erste Ansätze im Rahmen von NMPC [45], die aber algorithmisch deutlich aufwendiger als die „normale“ NMPC sind), oder aber d) durch die Integration höherer Ableitungen in der Optimierungsroutine. Die Sensitivität der Lösung nach unsicheren Parametern lässt sich formal als Ableitung berechnen. Moderne Ansätze in der robusten Optimierung erweitern daher das mathematische Modell oftmals um Terme mit höheren Ableitungen der Zielfunktion bzw. der Beschränkungen nach diesen Parametern. Trotz erfreulicher Fortschritte in diesem Bereich werden Unsicherheiten und die Bestimmung robuster Lösungen weiterhin eine große Herausforderung bleiben.

4.3 Weitere Herausforderungen

Während die Theorie und Algorithmik der Optimalen Steuerung und Regelung für dynamische Systeme entwickelt wurde, hat sich durch algorithmische Fortschritte und die Verfügbarkeit von schnellerer Hardware nach dem Mooreschen Gesetz die Tür geöffnet für die Betrachtung von **räumlichen Effekten**. Hiermit ist gemeint, dass die Temperatur in einem Raum nicht mehr anhand eines repräsentativen Messwertes an einer speziellen Stelle geregelt werden soll, sondern das Verhalten an jeder Stelle im Raum relevant wird.

Während dies für einfache Heizungen im Normalfall kaum nötig sein wird, gibt es viele Prozesse, bei denen die räumliche Komponente wichtig ist. Bei der Produktion von Stahlträgern beispielsweise ist strengstens darauf zu achten, dass die Temperaturunterschiede innerhalb des Trägers in der

Abkühlphase nicht zu groß werden, um spätere Risse zu vermeiden, [60]. Im Bereich der Aerodynamik spielt das Fluidverhalten eine entscheidende Rolle. An welchen Stellen im Raum sich Luftwirbel bilden und wo die Schockfronten verlaufen, beeinflusst direkt die Flugeigenschaften, [35].

Die Optimierung von Prozessen, die durch partielle Differentialgleichungen beschrieben werden, hat in den letzten Jahren einen großen Zuwachs an Aufmerksamkeit erfahren, beispielsweise durch ein Schwerpunktprogramm der Deutschen Forschungsgemeinschaft [15]. Trotzdem bleibt festzuhalten, dass die Kluft zwischen mathematischer Grundlagenforschung und praktischem Einsatz an dieser Stelle noch besonders groß zu sein scheint, nicht nur im Bereich der Algorithmen zur Lösung von Optimalsteuerungsproblemen, sondern auch in den systemtheoretischen Grundlagen der MPC-Methodik.

Eine Herausforderung ganz anderer Art stellt die Kombination von kontinuierlichen und diskreten Ereignissen dar. Mit kontinuierlich ist hier ein zusammenhängender Bereich gemeint, aus dem eine Variable oder eine Steuerung einen Wert annehmen kann. Man denke beispielsweise an die Geschwindigkeit eines Fahrzeugs oder die Position des Gaspedals. Mit diskret ist ein gerade nicht zusammenhängender Bereich gemeint, der mathematisch aus einer endlichen Menge möglicher Werte besteht. Hier ist die Gangschaltung ein gutes Beispiel für eine diskrete Steuerung, während „der Laufroboter hat Bodenkontakt oder nicht“ ein Zustand ist, bei dem entweder das eine oder das andere zutrifft. Ebenso führen digitale Kopplungen in vernetzten Systemen, die aus einer Vielzahl kleinerer physikalischer Teilsysteme bestehen, auf diskrete Zustandskomponenten. Systeme, die beide Typen von Zuständen und/oder Steuerungen vereinen, bezeichnet man als **hybride Systeme**. Oftmals ergeben sich hybride Systeme durch eine Mehrskalen-Modellierung. Hier werden schnelle transiente Übergänge als instantan angenommen. Ein gutes Beispiel hierfür sind Ventile, die nur offen oder geschlossen sein können, in der Praxis aber von einem Zustand in den anderen überführt werden müssen.

Mathematisch stellen die ganzzahlig oder diskret genannten Variablen, die sich bei der Modellierung ergeben, eine erhebliche Erschwerung der Optimierung dar. Dies mag auf den ersten Blick überraschen, stehen doch nur begrenzt viele Möglichkeiten zur Verfügung, die man theoretisch alle ausprobieren kann. Doch die Anzahl der Möglichkeiten wächst enorm schnell, wenn sich die Anzahl der Variablen erhöht. Dies ist bei hybriden Systemen der Fall, insbesondere um genau entscheiden zu können, wann und wie oft umgeschaltet werden soll. Eine Behandlung dieser Variablen durch Ausprobieren scheidet also wegen der immensen Anzahl an Möglichkeiten aus. Eine

Übersicht möglicher Methoden für ganzzahlige Steuerfunktionen und weiterführende Literatur zu hybriden Systemen findet man beispielsweise in [53].

Auch dieses Forschungsgebiet stellt sich als extrem aktiv dar, und wird beispielsweise durch das europäische Forschungsnetzwerk HYCON [31] und – für den Spezialfall digital vernetzter Systeme – durch einen DFG Forschungsschwerpunkt [16] gefördert. Insbesondere das Entwickeln verlässlicher und effizienter Algorithmen und eine Stabilitätsanalyse im Sinne des Abschnittes 3.3 werden in den nächsten Jahren viel Aufmerksamkeit erfordern, um das große Optimierungspotential bei hybriden Systemen ausschöpfen zu können.

5 Visionen und Handlungsempfehlungen

MPC und insbesondere NMPC gehören nach unserer Überzeugung zu den vielversprechendsten Methoden zur Lösung komplexer nichtlinearer Regelungsprobleme, weswegen sie in Zukunft weiter an Bedeutung gewinnen werden. Ihre Fähigkeit, Regelungsaufgaben nicht „nur“ lösen zu können, sondern über Optimalitätskriterien und Nebenbedingungen auch technische und ökonomische Beschränkungen einbeziehen zu können, liefert – gerade in Zeiten knapper werdender Ressourcen – deutliche Vorteile gegenüber anderen Methoden. Der rasante Fortschritt im Bereich der Optimierungsalgorithmen erlaubt dabei fortlaufend neue Anwendungen mit zunehmend schnellerer Dynamik.

Ganz im Sinne der Erfolgsfaktoren der klassischen Methoden von Hurwitz und Pontrjagin hat MPC darüberhinaus den entscheidenden Vorteil, intuitiv leicht verständlich zu sein und dadurch auch ohne tieferes Verständnis der dahinterstehenden Mathematik – seien es die systemtheoretische Analyse des Regelverhaltens oder die aus Numerik und Optimierung stammenden Grundlagen der verwendeten Algorithmen – anwendbar zu sein.

Wir wollen die Ausführungen im Folgenden daher auf diese Regelungsmethode beschränken, wenngleich sinngemäße Aussagen sicherlich auch für andere moderne Verfahren in der Steuerungs- und Regelungstechnik gelten.

5.1 Theorie und Praxis

Der aus theoretischer Sicht größte konzeptionelle Vorteil von (N)MPC – nämlich die explizite Einbeziehung eines mathematischen Modells zur Vorhersage des Systemverhaltens – ist zugleich einer der wesentlichen Nachteile in der Praxis, denn MPC verlangt dadurch zwingend nach hinreichend genauen mathematischen Modellen für die zu regelnden komplexen Prozesse.

Wie in Abschnitt 4.1 ausgeführt, gibt es hier inzwischen auch im nichtlinearen Bereich vielversprechende mathematische Ansätze. Trotzdem wird das Erstellen geeigneter Modelle aller Voraussicht nach zumindest im nichtlinearen Bereich vorerst aufwendig bleiben, so dass die dafür notwendigen Investitionen nur dann von der Industrie getätigt werden dürften, wenn die erzielbaren Vorteile zur Optimierung der Produktionsprozesse durch konkrete Beispiele belegt werden. Vielen der in Abschnitt 3.4 aufgeführten neueren Anwendungen kommt daher Pilotcharakter zu, deren Erfolg oder Misserfolg (und natürlich auch die Kommunikation entsprechender Erfolge) mittelfristig über die Akzeptanz speziell von NMPC in der industriellen Anwendung mitentscheiden wird.

Aber auch auf der theoretischen Seite gibt es noch eine Reihe offener Fragen. Insbesondere unterscheiden sich, wie in Abschnitt 3.3 ausgeführt, die Annahmen im größeren Teil der theoretischen MPC-Literatur deutlich von den in der Praxis verwendeten Verfahren. Hier erscheint es wünschenswert, diese Lücke zwischen Theorie und Praxis zu schließen. Gerade theoretische Arbeiten, die auf konkrete Verbesserungen von MPC-Verfahren abzielen – sei es nun verbesserte Stabilität, schnellere Algorithmen, bessere Regelgüte, höhere Robustheit – sollten versuchen, ihre Ergebnisse und Verbesserungen unter realistischen Annahmen zu erzielen.

Auf der algorithmischen Seite schließlich ist weiterer Fortschritt vor allem durch verstärkte Kooperationen von System- und Regelungstheoretikern auf der einen Seite und Optimierern und Numerikern auf der anderen Seite zu erhoffen. So kann z.B. nur eine systemtheoretische Stabilitätsanalyse des gesamten Regelalgorithmus Aufschluss darüber geben, in welchem Sinne die in Abschnitt 4.2 erläuterte Integration höherer Ableitungen in der Optimierungsroutine die Robustheit tatsächlich erhöht – und wie genau das Optimalitätskriterium dafür aussehen sollte. Auf der anderen Seite sollte z.B. die systemtheoretische Analyse von MPC-Algorithmen stets die Eigenheiten (und die in der Praxis vielfältig angewendeten „Tricks“) der Optimierungsroutinen berücksichtigen, um verlässliche Aussagen über die Effizienz der Regelung treffen zu können. In beiden Beispielen ist also die Kompetenz aus beiden Gebieten nötig.

5.2 Interdisziplinarität in der Ausbildung

All diese Ziele sind unseres Erachtens nur dann zu erreichen, wenn – ganz im Sinne der Zusammenarbeit von Hurwitz und Stodola – bereits in der Ausbildung Wert auf die Kooperation zwischen Mathematikern und Ingenieuren und zwischen den verschiedenen Gebieten der Angewandten Mathematik ge-

legt wird. Nur so können Mathematiker herausfinden, welche methodischen Herausforderungen die Praxisprobleme liefern, nur so können Ingenieure von den neuesten mathematischen Methoden profitieren und nur so können sich auch innerhalb der Mathematik Verfahren und Techniken aus verschiedenen Teildisziplinen sinnvoll ergänzen.

Erfreulicherweise gibt es hier eine ganze Reihe sehr vielversprechender aktueller interdisziplinärer Projekte, wie z.B.

- das Internationale Doktorandenkolleg „Identification, Optimization and Control with Applications in Modern Technologies“ der Universitäten Bayreuth, Erlangen-Nürnberg und Würzburg im Elitenetzwerk Bayern
- die Graduiertenschule „Mathematical and Computational Methods for the Sciences“ der Universität Heidelberg
- das Center of Excellence „Optimization in Engineering“ (OPTEC) an der Universität Leuven
- das Exzellenz-Cluster „Simulation Technology“ (SimTech) der Universität Stuttgart mit Doktorandenschule sowie Bachelor- und Master-Studiengängen

um nur diejenigen universitären Einrichtungen zu erwähnen, an denen die Autoren dieses Beitrags beteiligt sind. Die formalen Rahmenbedingungen für weitere Fortschritte sind derzeit gerade in den Bereichen Steuerungs- und Regelungstechnik also durchaus günstig. Es bleibt zu hoffen, dass diese Anstrengungen fortgesetzt werden, damit viele der dort entwickelten mathematischen Verfahren den Sprung in die industrielle Anwendung schaffen und die vorhandenen Chancen des Produktionsfaktors Mathematik auch in Zukunft genutzt werden.

Literatur

[1] BASF SE. BASF und Universität Heidelberg entwickeln gemeinsam neue Mathematik-Software für die Forschung. Press release P-08-308, June, 16th 2008.

[2] S. Bennett. *A history of control engineering 1800–1930*. Peter Peregrinus Ltd., London, 1979. Paperback reprint 1986.

[3] L.T. Biegler. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering*, 8:243–248, 1984.

[4] T. Binder, L. Blank, H.G. Bock, R. Bulirsch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J.P. Schlöder, and O.v. Stryk. Introduction to model based optimization of chemical processes on moving horizons. In M. Grötschel, S.O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems: State of the Art*, pages 295–340. Springer, 2001.

[5] C. C. Bissell. Stodola, Hurwitz and the genesis of the stability criterion. *Int. J. Control*, 50:2313–2332, 1989.

[6] H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics. In K.H. Ebert, P. Deuflhard, and W. Jäger, editors, *Modelling of Chemical Reaction Systems*, volume 18 of *Springer Series in Chemical Physics*, pages 102–125. Springer, Heidelberg, 1981.

[7] H.G. Bock. Recent advances in parameter identification techniques for ODE. In P. Deuflhard and E. Hairer, editors, *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, pages 95–121. Birkhäuser, Boston, 1983.

[8] V. G. Boltyanski, R. V. Gamkrelidze, and L. S. Pontryagin. On the theory of optimal processes (russian). *Dokl. Akad. Nauk SSSR*, 110:7–10, 1956.

[9] C.I. Byrnes, A. Isidori, and J.C. Willems. Passivity, feedback equivalence, and the global stabilization of minimum phase nonlinear systems. *IEEE Transactions on Automatic Control*, 36(11):1228–1240, 1991.

[10] J.-B. Caillau, J. Gergaud, T. Haberkorn, P. Martinon, and J. Noailles. Numerical optimal control and orbital transfers. In *Proceedings of the Workshop Optimal Control, Sonderforschungsbreich 255: Transatmosphärische Flugsysteme, Heronymus Munchen, ISBN 3-8979-316-X*, pages 39–49, Greifswald, Germany, 2002.

[11] E. F. Camacho and C. Bordons. *Model predictive control*. Springer–Verlag, London, 2nd edition, 2004.

[12] M. Canale, L. Fagiano, M. Ippolito, and M. Milanese. Control of tethered airfoils for a new class of wind energy generators. In *Proceedings of the 45th IEEE Conference on Decision and Control, San Diego, California*, pages 4020–4026, 2006.

[13] H. Chen and F. Allgöwer. A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability. *Automatica*, 34(10):1205–1217, 1998.

[14] M.A. Dahleh and I.J. Diaz-Bobillo. *Control of Uncertain Systems : A Linear Programming Approach*. Prentice Hall, Englewood Cliffs, N.J, 1995.

[15] DFG Schwerpunktprogramm 1253, Optimization with Partial Differential Equations. http://www.am.uni-erlangen.de/home/spp1253/.

[16] DFG Schwerpunktprogramm 1305, Regelungstheorie digital vernetzter dynamischer Systeme. http://spp-1305.atp.rub.de/.

[17] M. Diehl, H.G. Bock, and J.P. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. *SIAM Journal on Control and Optimization*, 43(5):1714–1736, 2005.

[18] M. Diehl, H.G. Bock, J.P. Schlöder, R. Findeisen, Z. Nagy, and F. Allgöwer. Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. *J. Proc. Contr.*, 12(4):577–585, 2002.

[19] M. Diehl, R. Findeisen, F. Allgöwer, H.G. Bock, and J.P. Schlöder. Nominal stability of the real-time iteration scheme for nonlinear model predictive control. *IEE Proc.-Control Theory Appl.*, 152(3):296–308, 2005.

[20] R. Dittmar and B.-M. Pfeiffer. Modellbasierte prädiktive Regelung in der industriellen Praxis. *at – Automatisierungstechnik*, 54:590–601, 2006.

[21] H.J. Ferreau, P. Ortner, P. Langthaler, L. del Re, and M. Diehl. Predictive control of a real-world diesel engine using an extended online active set strategy. *Annual Reviews in Control*, 31(2):293–301, 2007.

[22] M. Fliess, J. Lévine, P. Martin, and P. Rouchon. Flatness and defect of non-linear systems: introductory theory and examples. *International Journal of Control*, 61(6):1327–1361, 1995.

[23] O. Föllinger. *Optimierung dynamischer Systeme: Eine Einführung für Ingenieure*. Oldenbourg, München, 2. edition, 1988.

[24] B. Francis, J. Helton, and G. Zames. H_∞-optimal feedback controllers for linear multivariable systems. *IEEE Transactions on Automatic Control*, 29(10):888–900, 1984.

[25] R. Franke, M. Meyer, and P. Terwiesch. Optimal control of the driving of trains. *at – Automatisierungstechnik*, 50(12):606–614, 2002.

[26] G. Grimm, M. J. Messina, S. E. Tuna, and A. R. Teel. Model predictive control: for want of a local control Lyapunov function, all is not lost. *IEEE Transactions on Automatic Control*, 50(5):546–558, 2005.

[27] L. Grüne. Computing stability and performance bounds for unconstrained NMPC schemes. In *Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, Louisiana*, pages 1263–1268, 2007.

[28] L. Grüne and A. Rantzer. On the infinite horizon performance of receding horizon controllers. *IEEE Transactions on Automatic Control*, 2008. To appear.

[29] G.A. Hicks and W.H. Ray. Approximation methods for optimal control systems. *Can. J. Chem. Engng.*, 49:522–528, 1971.

[30] A. Hurwitz. Über die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt. *Math. Annalen*, 46:273–284, 1895. Nachgedruckt in: R. Jeltsch (ed.) et al., Stability theory. Birkhäuser, Basel, 1996, 239–249.

[31] HYCON, European Network of Excellence on Hybrid Control. http://www.ist-hycon.org/.

[32] A. Ilzhoefer, B. Houska, and M. Diehl. Nonlinear MPC of kites under varying wind conditions for a new class of large scale wind power generators. *International Journal of Robust and Nonlinear Control*, 17(17):1590–1599, 2007.

[33] A. Isidori. *Nonlinear control systems*, volume 1. Springer, Berlin, 3. edition, 2002.

[34] A. Jadbabaie and J. Hauser. On the stability of receding horizon control with a general terminal cost. *IEEE Transactions on Automatic Control*, 50(5):674–678, 2005.

[35] A. Jameson. Aerodynamics. In E. Stein, R. De Borst, and T.J.R. Hughes, editors, *Encyclopedia of Computational Mechanics*, volume 3, pages 325–406. Wiley, 2004.

[36] R. E. Kalman. When is a linear control system optimal? *Trans. ASME, Series D, Journal of Basic Engn.*, 86:51–60, 1964.

[37] I. Kanellakopoulos, P.V. Kokotovic, and A.S. Morse. Systematic design of adaptive controllers for feedback linearizable systems. *IEEE Transactions on Automatic Control*, 36(11):1241–1253, 1991.

[38] S. S. Keerthy and E. G. Gilbert. Optimal infinite horizon feedback laws for a general class of constrained discrete-time systems: stability and moving horizon approximations. *J. Optimiz. Theory Appl.*, 57:265–293, 1988.

[39] S. Körkel. *Numerische Methoden für Optimale Versuchsplanungsprobleme bei nichtlinearen DAE-Modellen*. PhD thesis, Universität Heidelberg, Heidelberg, 2002.

[40] S. Körkel, E. Kostina, H.G. Bock, and J.P. Schlöder. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optimization Methods and Software*, 19:327–338, 2004.

[41] M. Krstic, I. Kanellakopoulos, and P.V. Kokotovic. *Nonlinear and Adaptive Control Design*. Wiley, New York, 1995.

[42] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems*. Wiley, New York, 1972.

[43] D.B. Leineweber, I. Bauer, A.A.S. Schäfer, H.G. Bock, and J.P. Schlöder. An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization (Parts I and II). *Computers and Chemical Engineering*, 27:157–174, 2003.

[44] W.C. Li, L.T. Biegler, C.G. Economou, and M. Morari. A constrained pseudo-Newton control strategy for nonlinear systems. *Computers and Chemical Engineering*, 14(4/5)(451-468), 1990.

[45] D. Limon, T. Alamo, F. Salas, and E. F. Camacho. Input to state stability of min-max MPC controllers for nonlinear systems with bounded uncertainties. *Automatica*, 42(5):797–803, 2006.

[46] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: stability and optimality. *Automatica*, 36:789–814, 2000.

[47] V. L. Mehrmann. *The autonomous linear quadratic control problem. Theory and numerical solution*, volume 163 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 1991.

[48] Z. Nagy, B Mahn, R. Franke, and F Allgöwer. Evaluation study of an efficient output feedback nonlinear model predictive control for temperature tracking in an industrial batch reactor. *Control Engineering Practice*, 15:839–850, 2007.

[49] H. J. Pesch. *Schlüsseltechnologie Mathematik*. Teubner, Stuttgart, 2002.

[50] H.J. Pesch and R. Bulirsch. The maximum principle, Bellman's equation and Caratheodory's work. *Journal of Optimization Theory and Applications*, 80(2):203–229, 1994.

[51] L.S. Pontryagin, V.G. Boltyanski, R.V. Gamkrelidze, and E.F. Miscenko. *The Mathematical Theory of Optimal Processes*. Wiley, Chichester, 1962.

[52] S. J. Qin and T. A. Badgwell. An overview of nonlinear model predictive control applications. In J. C. Cantor, C. E. Garcia, and B. Carnahan, editors, *Nonlinear model predictive control*. Birkhäuser, Basel, 2000.

[53] S. Sager, G. Reinelt, and H.G. Bock. Direct methods with maximal lower bound for mixed-integer optimal control problems. *Mathematical Programming*, published online at http://dx.doi.org/10.1007/s10107-007-0185-6 on 14 August 2007, 2008.

[54] R.W.H. Sargent and G.R. Sullivan. The development of an efficient optimal control package. In J. Stoer, editor, *Proceedings of the 8th IFIP Conference on Optimization Techniques (1977), Part 2*, Heidelberg, 1978. Springer.

[55] A.J. van der Schaft. *L_2-gain and Passivity Techniques in Nonlinear Control*. Springer, London, 2. edition, 2000.

[56] V.H. Schulz. Solving discretized optimization problems by partially reduced SQP methods. *Computing and Visualization in Science*, 1:83–96, 1998.

[57] Y. Shimizu, T. Ohtsuka, and M. Diehl. A real-time algorithm for nonlinear receding horizon control using multiple shooting and continuation/krylov method. *International Journal of Robust and Nonlinear Control*, 2008. (in print).

[58] H. Sira-Ramírez and S.K. Agrawal. *Differentially Flat Systems*. Marcel Dekker, New York, 2004.

[59] B. Srinivasan, S. Palanki, and D. Bonvin. Dynamic Optimization of Batch Processes: I. Characterization of the nominal solution. *Computers and Chemical Engineering*, 27:1–26, 2003.

[60] F. Tröltzsch and A. Unger. Fast solution of optimal control problems in the selective cooling of steel. *Zeitschrift für Angewandte Mathematik und Mechanik*, pages 447–456, 2001.

[61] T.H. Tsang, D.M. Himmelblau, and T.F. Edgar. Optimal control via collocation and non-linear programming. *International Journal on Control*, 21:763–768, 1975.

[62] J. Wischnegradski. Sur la théorie générale des régulateurs. *Comptes Rendus de L'Académie des Sciences de Paris*, 83:318–321, 1876.

[63] V. M. Zavala and L.T. Biegler. The advanced step NMPC controller: Optimality, stability and robustness. *Automatica*, 2008. (accepted for publication).

[64] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, Upper Saddle River, NJ, 1996.