INAUGURAL-DISSERTATION
zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht-Karls-Universität Heidelberg

vorgelegt von
M.Sc. Huu Chuong La
aus Ninh Binh

Tag der mündlichen Prüfung:

Dual Control for Nonlinear Model Predictive Control

Gutachter: Prof. Dr. Dr. h.c. mult. Hans Georg Bock

Zusammenfassung

Diese Arbeit behandelt das Dual Control Problem für nichtlineare modellprädiktive Regelung (Nonlinear Model Predictive Control, kurz NMPC) aus der Perspektive der optimalen Versuchsplanung (Optimal Experimental Design, kurz OED).

In den letzten Jahren stellt die Steuerung unsicherer Prozesse Mathematiker vor große Herausforderungen, bietet aber auch Chancen. Obwohl das Prinzip der dynamischen Programmierung gelten könnte, ist seine Anwendbarkeit auf sehr einfache Fälle beschränkt. Dies erfordert das Studium von Approximationsmethoden sowie effizienter Echtzeit-Algorithmen.

In dieser Arbeit untersuchen wir Optimalsteuerungsprobleme mit unsicheren Parametern und Zuständen und stellen neue auf Dual Control basierende Methoden vor. Wir führen zunächst eine Sensitivitätsanalyse durch, um die Auswirkung der Unsicherheit auf die Güte der Steuerung zu bewerten. Ausgehend von der Analyse des Zusammenspiels zwischen der Optimalsteuerungsaufgabe und der Aufgabe des Informationsgewinns schlagen wir neue Ansätze im NMPC-Kontext mithilfe von OED vor. Hierzu präsentieren wir den statistischen Hintergrund und probabilistische Schranken für die eigentliche Controller-Performance bezüglich der ursprünglichen Zielfunktion. Damit füllen wir eine Lücke in der Literatur.

Wenn NMPC den Prozess im Laufe der Zeit steuert und das Schätzverfahren läuft, ist es von Interesse, die Konvergenzeigenschaften und das asymptotische Verhalten zu verstehen. Wir widmen der Untersuchung einiger Eigenschaften von Least-Squares (LS)-Schätzern ein Kapitel. Es wird gezeigt, dass das Schätzproblem in manchen Fällen schlecht gestellt ist. Dies führt zu Divergenz. Jedoch kann Konvergenz bei Verwendung einer sequentiellen LS-Methode beibehalten werden. Ein Konvergenzresultat wird hergeleitet.

Andererseits haben wir beobachtet, dass die Konvergenz der Parameterschätzung in einigen Prozessen irrelevant werden kann, wenn die wichtigen Zustände stabil sind. Dies motiviert unsere Untersuchung der Teilstabilität für NMPC, die mehrere Grundergebnisse der klassischen Stabilitätsanalyse für NMPC erweitert.

Eine weitere Motivationsquelle dieser Arbeit ist die Untersuchung der nichtlinearen optimalen Versuchsplanung. Weil optimale Designs für bestimmte Werte der Parameter berechnet werden, deren wahre Werte aber nicht bekannt sind, ist es wichtig, die Optimalität der Designs zu beurteilen sowie Verfahren zu finden, um die wahren Parameterwerte zu erreichen. Dies motiviert die Erforschung von sequentiellem OED im Rahmen des Dual Control. Hierzu formulieren wir das sequentielle OED Problem, sodass das Prinzip der dynamischen Programmierung anwendbar ist. Wir beweisen einige Resultate über Designs mit endlichen Trägern und bauen so eine Brücke zwischen kontinuierlichen Designs und diskreten Designs.

Die Methoden wurden im Rahmen dieser Arbeit implementiert. Wir illustrieren die erzielten Ergebnisse, indem wir die Aufgaben behandeln, die von klassischen Beispielen aus der Literatur bis hin zu praktischen Anwendungen in der Fahrzeugsteuerung und in der chemischen Verfahrenstechnik reichen.

Abstract

In this thesis we treat the problem of Dual Control for Nonlinear Model Predictive Control (NMPC) from a perspective of Optimal Experimental Design (OED).

Controlling uncertain processes poses great challenges as well as offers opportunities for mathematicians in recent years. While the Dynamic Programming principle might hold, its applicability is limited to a few very simple cases. This calls for the study of approximation methods and real-time algorithms.

In this work we study optimal control problems with uncertain parameters and states and develop new methods based on Dual Control. We first carry out a *sensitivity analysis* to assess the effect of uncertainty on the control performance. By analyzing the interplay between the performance control task and the information gain task, we propose novel approaches to the Dual Control problem in the context of NMPC with the help of OED. Furthermore, we present the statistical background and probabilistic bounds for the realized controller performance with respect to the original objective. Therefore, we essentially fill a gap in the literature.

As NMPC drives the process in the course of time and the estimation procedure runs, it is of interest to understand the convergence properties and the asymptotic properties of the parameter and state estimates. We devote one chapter to the investigation of *asymptotic properties* of the Least Squares (LS) estimates, showing that in some cases the estimation problem is ill-posed leading to divergence. With the use of a sequential LS method, however, convergence can be retained. A convergence result is established.

On the other hand, we observe that for some processes, if the states of interest are stable, the convergence of parameter estimates may become irrelevant. This motivates our study on *partial stability for NMPC* which extends the classical stability analysis of NMPC by several fundamental results, including general stability results without terminal costs or terminal constraints.

Another source of motivation for this thesis is the study of nonlinear Optimal Experimental Design. Since optimal designs are computed for specific values of parameters but the true ones are unknown, it is important to assess the optimality of designs as well as to find a way to reach the true parameters. This motivates our study on *sequential OED* in the framework of Dual Control. For this purpose, we reformulate the problem of sequential OED to make it applicable for the Dynamic Programming principle. We also build a bridge between continuous designs and discrete designs by presenting several results on finite support for continuous OED.

The methods have been implemented and we illustrate the obtained results by examples ranging from classics to practical applications in vehicle control and chemical engineering.

Danksagungen

Ich möchte mich zutiefst bei Prof. Hans Georg Bock (Fakultät für Mathematik und Informatik, Universität Heidelberg) für seine intensive Betreuung und Unterstützung bedanken. Dieser Dank geht auch an Dr. Andreas Potschka und Dr. Johannes Schlöder. Durch Ihre Ideen, Anregungen und Korrekturen wurde diese Arbeit Schritt für Schritt vorangebracht. Weiterhin habe ich unschätzbare Erfahrungen gesammelt, die mein wissenschaftliches Leben erheblich geprägt haben. Von Herzen danke ich Ihnen für Ihre Hilfe und Geduld.

Die wunderbare Atmosphäre am interdisziplinären Zentrum für wissenschaftliches Rechnen (IWR) spielte eine entscheidende Rolle in meiner Arbeit. Bei verschiedenen Problemen bezüglich Arbeit und Leben habe ich nützliche Hilfe bekommen. An alle Kollegen und Kolleginnen, insbesondere die Mitglieder der MOBOCON Gruppe, geht mein herzlicher Dank für Ihre Freundschaft und Hilfe.

Ich danke Prof. Donald Richards (Pennsylvania State University) für die Zusammenarbeit bei den asymptotischen Eigenschaften des LQ-Verfahrens und die gründliche Korrektur des Kapitels 3. Diskussionen mit Ihm sind immer unterhaltsam und haben mich sehr motiviert.

Meine Promotion wurde am Anfang von dem Erasmus Mundus Mobility with Asia (EMMA) und danach von der Heidelberger Graduiertenschule der mathematischen und computergestützten Methoden für die Wissenschaften (HGS MathComp) gefördert. Der HGS danke ich für die wissenschaftliche Unterstützung durch zahlreiche akademische Aktivitäten und Workshops.

Dem Mathematischen Institut von Hanoi danke ich dafür, dass mir das Studium in Heidelberg ermöglicht wurde. An dieser Stelle möchte ich mich bei Prof. Dinh Nho Hao, Prof. Ngo Viet Trung und Prof. Hoang Xuan Phu für Ihre Unterstützung bedanken.

Schließlich danke ich meinen Eltern, meinem Bruder und meinen Schwestern, die mich fortwährend unterstützen.

Contents

List of acronyms

DC	Dual Control
DP	Dynamic Programming
EKF	Extended Kalman Filter
HJBE	Hamilton-Jacobi-Bellman Equation
IVP	Initial Value Problem
KF	Kalman Filter
LQG	Linear Quadratic Gaussian
LS	Least Squares
MHE	Moving Horizon Estimation
MLE	Maximum Likelihood Estimator
NMPC	Nonlinear Model Predictive Control
OCP	Optimal Control Problem
ODE	Ordinary Differential Equation
OED	Optimal Experimental Design
PMP	Pontryagin Minimum Principle

Introduction

Control of uncertain processes has spread over many mathematical research areas in recent years and has found remarkable applications in chemical engineering, vehicle control, robotics, etc. Having feedback in nature, Nonlinear Model Predictive Control (NMPC) stands out as a popular and efficient control methodology to deal with uncertainty (see e.g., Grüne and Pannek [36], Rawlings and Mayne [71], Camacho and Bordons [16]). Nominal NMPC, in which the *Certainty Equivalence* principle is implicitly assumed to hold, considers the estimates of parameters and states as if they were the true values. Corrections and adjustments are made at each NMPC step with the help of an estimation procedure such as extended Kalman filters (EKF) or Moving Horizon Estimation (MHE) (see Rawlings and Mayne [71]). Nominal NMPC performs well in many but not all cases. Nonrobust behavior of control and estimation can happen, even instability and infeasibility might occur. Also severe loss of optimality is often the case. This is because nominal NMPC ignores the discrepancy between the estimates and the true values and leaves out the possibility to increase the accuracy of the future estimates. The future estimates when new measurements arrive may not be reliable enough, resulting in a big mismatch between prediction and reality. It is therefore of interest in both theory and practice to surmount the weaknesses of nominal NMPC and develop more powerful control methods.

Dual Control refers to strategies that balance control and estimation. The concept of Dual Control first appeared in the 1960s with the pioneering work of Feldbaum [26]. Further analysis and clarification have been carried out, for example in Åström [4], Wittenmark [86], Filatov and Unbehauen [27]. The problem of Dual Control was approached by stochastic Dynamic Programming. However, this approach has experienced impediments in practical applications due to the "curse of dimensionality" and complications in computing conditional expectations. This calls for reformulations and approximation methods. Certainty Equivalence control, also known as nominal control, provides a handy approximation method. Relying on the Certainty Equivalence principle, however, it ignores the uncertainties in the current estimates, and hence may not be desirable. Dual NMPC is an improved variant of NMPC that strikes a balance between the goal of control and the goal of estimation. It takes the inexactness of the current estimates and the possibility of enhancing their accuracy into account when solving an Optimal Control Problem (OCP) at each step. With more accurate estimates, better control actions can be obtained in the future, leading to a good overall performance.

In the literature, Dual Control is often tackled heuristically. The objective function is penalized by some scalarization of the covariance matrix with some weight, see Wittenmark [86], Filatov and Unbehauen [27]. The use of OED has not been thoroughly explored. Numerical methods have not been developed adequately, especially for nonlinear processes modeled by systems of differential equations. Moreover, NMPC

is a natural framework for Dual Control but the combination has not been investigated comprehensively. Very recent studies explored the use of OED. For example, Heirung et al. [37] constructed an approximation method for linear input-output models; Lucia and Paulen [58] presented an approach to Dual Control for robust NMPC based on scenario trees, in which parameters are supposed to assume only a finite number of values. The computational efforts for such approaches are tremendous. Above all there is still a big gap in the literature on the theoretical background of approximation methods for Dual Control. As a consequence, the applicability of Dual Control in real-life applications still needs to be justified. Therefore, our goal is to offer a more rigorous and viable treatment of Dual Control for complex processes together with efficient real-time numerical methods.

Contributions of the thesis

In this thesis, we treat OCPs under uncertainty using an NMPC strategy. From the OED point of view, we underline the two tasks that the control should care about: performance control task and information gain task. The former aims to operate the process feasibly and in an optimal way specified by the objective function. The latter aims to gain information about the process in order to get reliable estimates. By analyzing the relationship between the two tasks, we gain insights into the behavior of nominal NMPC, explaining why it works well in some cases and why it yields unsatisfactory results in other cases. We then propose novel Dual Control approaches in the context of NMPC. The covariance matrix of the future estimates is weighted by the sensitivities of the optimal nominal objective value with respect to uncertain parameters and states. This quantity can be interpreted as the predictive variance of the optimal nominal objective value. By reducing this variance, we can improve the accuracy of the future estimates of parameters and states and increase the probability of state constraint fulfillment. On the one hand, our methods are supported by stochastic optimization. On the other hand, it aims to balance the performance control task and the information gain task. Moreover, we provide the statistical background for our approaches and probabilistic bounds for the controller performance with respect to the original objective, which are missing in the literature. As soon as the parameter estimates are good enough, i.e., the covariance matrix is small or the parameters have little influence on the objective function, i.e., the sensitivity is weak, the dual effect would be small. We can adaptively switch to nominal NMPC in order to save computational efforts. The weaknesses of nominal NMPC and the potential of dual NMPC are illustrated by a collection of examples in vehicle control and chemical engineering. Furthermore, this thesis investigates some interesting issues related to NMPC such as sequential OED, convergence of least squares (LS) estimates, partial stability for NMPC, providing insights into controlling uncertain processes. A summary of the main contributions of the thesis is as follows.

- New approaches to dual NMPC based on OED and sensitivity analysis are proposed together with the theoretical background. Corresponding numerical methods are developed and implemented. The resulting software runs fast and yields consistent results on various test problems as well as complex processes in chemical engineering.

- A comprehensive study of the interplay between the performance control task and the information gain task for controlling uncertain processes illustrates concrete

cases when they are conflicting and when they align.

- A study on partial stability of NMPC together with new results that extend classical results in the stability theory of nominal NMPC are given. Most notably, we establish partial stability of NMPC without using terminal constraints and terminal costs. This is the state-of-the-art setup for NMPC stability analysis. It is also shown that for some processes, some states are asymptotically stable and have little effect on the objective function. For such processes, nominal NMPC often performs well.

- An analysis of some asymptotic properties of the LS estimates in connection with MHE is carried out. It is shown that the LS objective function for large sample sizes may be nearly independent of the estimated parameters, making the LS problem ill-posed. To overcome this difficulty, we propose a sequential LS strategy and establish convergence results.

- An investigation of OED for nonlinear systems in the continuous case is carried out. We prove that every design is equivalent to a design with finite support. Moreover for optimal designs, sharp upper and lower bounds for the number of support points are provided.

- A formulation of sequential OED for nonlinear systems is proposed and solved by Dual Control.

Thesis overview

The thesis is organized as follows. Chapter 1 presents a general introduction to NMPC with highlights on the Certainty Equivalence principle and Dual Control for NMPC. Chapter 2 gives a problem-solving-oriented introduction to OCPs including the Pontryagin Minimum Principle (PMP) and the Hamilton-Jacobi-Bellman Equation (HJBE). We carry out a comprehensive sensitivity analysis of OCPs under uncertainty based on both PMP and HJBE, revealing their similarities and differences. Asymptotic properties of the LS estimators and convergence analysis of sequential LS are presented in Chapter 3. Chapter 4 deals with nonlinear OED in the continuous case. Dual Control begins in Chapter 5 with an introduction to the theory, its connection with NMPC and a solution to sequential OED based on Dual Control. Chapter 6 is devoted to new approaches to dual NMPC based on OED. We propose corresponding algorithms together with numerical methods. Chapter 8 illustrates the behavior of dual NMPC in a collection of examples. There we give a thorough assessment of the performance of dual NMPC and nominal NMPC. Furthermore, the interplay between performance control and information gain is elucidated. It is observed that dual NMPC performs better and more robust than nominal NMPC and has potential in controlling processes with limited number of sensors. We also point out that for some parameters, nominal NMPC can perform well for estimating them, dispensing the need for Dual Control. Chapter 7 sheds light on this behavior in which partial stability for NMPC is studied. Two appendices deliver proofs of some theoretical results, including the derivation of HJBE, Kalman filter and some properties of linear estimation. We conclude the thesis with an outlook on future research.

A part of the thesis results has been published or submitted for publication. The contents of Chapter 4 are based on La et al. [55]. The material of Section 6.4 and Chapter 8 can be found in La et al. [53, 54]. Chapter 7 presents the results of La et al. [52].

For convenience, we introduce the notation that we use throughout this thesis. If A is a symmetric matrix of dimension $n \times n$ then the notation $A \succ 0$ (respectively, $A \succeq 0$) means that A is positive definite (respectively, positive semidefinite). Similarly, for symmetric matrices $A,\ B \in \mathbb{R}^{n\times n}$, $A \succ B$ (respectively, $A \succeq B$) means that $A - B \succ 0$ (respectively, $A - B \succeq 0$). The identity matrix of dimension $n \times n$ is denoted by $\mathbb{I}_n$. For a random vector X, we denote by $\mathbb{E}X$ the expectation and by $\mathrm{Cov}(X)$ the covariance of X. Furthermore, the notation $\mathcal{N}(x_0, R)$, where $x_0 \in \mathbb{R}^n$ and $R \in \mathbb{R}^{n\times n}$, $R \succeq 0$, stands for the Gaussian distribution with mean x_0 and covariance matrix R. By $X \sim \mathcal{N}(x_0, R)$ we understand that X is a random vector having distribution $\mathcal{N}(x_0, R)$. Finally, we denote by $\mathbb{R}^+$ the set of all nonnegative real numbers.

Chapter 1

Preliminaries

1.1 Optimal Control Problems under uncertainty

Optimal control has enjoyed exciting developments since the 1950s. With advances in both theory and numerical methods, optimal control has become more and more widespread in various areas of application such as chemical engineering, automobile, robotics, aerospace, etc. From a theoretical point of view, it is an extension of the variational calculus initiated by Euler and Lagrange. The corner stones include the Bellman Dynamic Programming Principle together with the Hamilton-Jacobi-Bellman Equation (HJBE) and the Pontryagin Minimum Principle (PMP). Furthermore, Kalman made seminal contributions to both the theory and computational methods for optimal control and parameter and state estimation, among which are the use of the state-space method, the linear quadratic regulator (LQR) and the Kalman filter.

Typically, an optimal control problem (OCP) comprises three components:

- A process to be controlled, which is usually modeled by a system of ordinary differential equations (ODEs), differential algebraic equations (DAEs), partial differential equations (PDEs) or difference equations.
- An objective function or performance index. It may be an economic goal or a deviation from set-points when doing tracking.
- A set of constraints including boundary conditions, constraints on states and controls. They can be safety requirements, positiveness of physical quantities, etc.

A rather general form of an OCP is

$$\min_{u(\cdot),x(\cdot),t_f} \Phi(x(t_f)) + \int_{t_0}^{t_f} L(t, x(t), u(t))dt \tag{1.1}$$

subject to

$$\dot{x}(t) = f(t, x(t), u(t), p), \quad t \in [t_0, t_f], \tag{1.2}$$
$$x(t_0) = x_0, \tag{1.3}$$
$$u(t) \in \mathbb{U}, \quad t \in [t_0, t_f], \tag{1.4}$$
$$\Psi(t_f, x(t_f)) = 0, \tag{1.5}$$

where $x \in \mathbb{R}^{n_x}$ is the vector of states, u the vector of controls. The set $\mathbb{U} \subseteq \mathbb{R}^{n_u}$, assumed to be closed, is used to enforce control constraints. Furthermore, $\Psi : \mathbb{R} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_\Psi}$ expresses terminal constraints

We note that $p \in \mathbb{R}^{n_p}$ models constant parameters. In real applications, the true values of p may not be available beforehand. Also there can be disturbances on the right-hand side due to imperfect models. This thesis mainly considers unknown constant parameters and uncertain initial states as uncertainties in the process. However, for linear systems such as Linear Quadratic Gaussian presented in Chapters 5–6, we are able to treat disturbances on the right hand side. Furthermore we pursue a soft constraint approach to deal with constraints, including path constraints and terminal constraints. Hence path constraints, and sometimes terminal constraints are absent from the problem formulation. In addition, when the parameters p are considered to be known, we will omit p for brevity. This omission also applies when we view parameters as auxiliary states for a concise presentation.

Though the final time t_f is allowed to vary, OCP (1.1)–(1.5) can be transformed into an OCP with fixed final time. This makes numerical integration of the ODEs for numerically solving OCPs possible.

Let $\tau \in [t_0, t_f]$ and $x^0 \in \mathbb{R}^{n_x}$. For given p, $u(\cdot)$, we assume that the solution of the initial value problem (IVP) (1.2) with the initial condition $x(\tau) = x^0$ exists and is unique, and denote it by $x(t; \tau, x^0, u(\cdot), p)$.

Discrete time problems are of interest in their own right, which are treated parallel with the continuous-time counterpart in Chapters 2, 5, 6. To point out the connection, we show how to transform the continuous time OCP (1.1)–(1.5) into a discrete time one. Consider a sampling grid $t_0 < t_1 < t_2 < ... < t_N = t_f$. The control $u(\cdot)$ is often parametrized on this grid in form $u(t) = \varphi_k(t, u_k)$ for $t \in [t_k, t_{k+1})$, where $u_k \in \bar{\mathbb{U}} \subseteq \mathbb{R}^{\nu_u}$ and $\varphi_k : \mathbb{R} \times \mathbb{R}^{\nu_u} \to \mathbb{U}$ can be constant, linear or quadratic in t. For $k \geq 0$, we define $x_{k+1} = x(t_{k+1}; t_k, x_k, \varphi_k(\cdot, u_k), p)$ and $f_k : \mathbb{R}^n \times \bar{\mathbb{U}} \to \mathbb{R}^n$ by $f_k(x_k, u_k) = x(t_{k+1}; t_k, x_k, \varphi_k(\cdot, u_k), p)$. Furthermore, we set

$$F_N(x_N) = \Phi(x_N), \quad L_k(x_k, u_k) = \int_{t_k}^{t_{k+1}} L(t, x(t), u(t))dt$$

with $x(t) = x(t; t_k, x_k, \varphi_k(\cdot, u_k), p)$. We then arrive at the discrete time OCP

$$\min_{u_0,...,u_{N-1}} F_N(x_N) + \sum_{k=0}^{N-1} L_k(x_k, u_k) \tag{1.6}$$

subject to

$$x_{k+1} = f_k(x_k, u_k), \quad k = 0, 1, 2, ..., N-1 \tag{1.7}$$

with x_0 given. Moreover the terminal constraint (1.5) can be included in the form $\Psi(t_N, x_N) = 0$.

Parallel with theoretical investigation, numerical methods for OCPs have been studied extensively. We can name two prominent approaches: indirect methods and direct methods. Indirect methods first derive optimality conditions using theoretical principles like PMP, HJBE, and then compute the optimal controls and corresponding states numerically by some discretization scheme. Direct methods first discretize controls and states, transform the problem into nonlinear programs, then use state-of-the-art numerical methods to solve the discretized problem. Among popular numerical methods,

multiple shooting together with well-developed supplementary techniques has proved to be efficient and successful in industrial applications, see Section 1.8.

Increasing complexity of nonlinear modeling, demands on computational speed for large scale problems and online operation, the wish for improving productivity and the need to ensure safe operation pose great challenges for researchers in the field. Moreover, uncertainties that are often present in real processes need to be coped with efficiently. Here arises the problem of estimation and the problem of the relationship between the quality of estimates and the control performance. This initiates the concept of *Dual Control* which constitutes the main subject of this thesis, see Section 1.10.

1.2 Parameter and state estimation

Often in real-life applications, the parameters of the process are not known exactly and need to be estimated. Moreover the states cannot be fully measured nor measured with perfect accuracy. It is therefore essential to estimate the current states and parameters in order to validate the model and control the process. In connection with OCP (1.1)–(1.5), we consider a measurement function $\eta : \mathbb{R} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ together with noisy measurements $y(t) \in \mathbb{R}^{n_y}$, i.e.,

$$y(t) = \eta(t, x(t; t_0, x_0, u(\cdot), p)) + \varepsilon(t)$$

where $\varepsilon(t)$ are the measurement noise, often assumed to be independent white noise. In reality, it is often the case that measurements are obtained by sampling at discrete time points, say $t_1, t_2, ...$ with

$$y(t_i) = \eta(t_i, x(t_i; t_0, x_0, u(\cdot), p)) + \varepsilon(t_i), \quad i = 1, 2, ...,$$

where $\varepsilon(t_i) \sim \mathcal{N}(0, R_i)$. For brevity, we set

$$x_i = x(t_i; t_0, x_0, u, p), \quad y_i = y(t_i), \quad i = 1, 2, ...,$$

and keep in mind that they all depend on (x_0, p). With u given, based on N measurements, the parameters and the initial states are estimated by minimizing some criterion function

$$\min_{x_0, p} Q(x_0, p)$$

subject to (1.2)–(1.3). A typical criterion function is the least squares (LS) function

$$Q(x_0, p) = \frac{1}{2} \sum_{i=1}^{N} (y_i - \eta(t_i, x_i))^T R_i^{-1} (y_i - \eta(t_i, x_i)). \tag{1.8}$$

If an a priori estimate $\overline{\nu}_0$ of $\nu_0 = (x_0, p)$ together with a nonsingular initial covariance matrix Σ_0 are available, we can use the regularized LS function

$$\bar{Q}(x_0, p) = \frac{1}{2} (\nu_0 - \overline{\nu}_0)^T \Sigma_0^{-1} (\nu_0 - \overline{\nu}_0) + \frac{1}{2} \sum_{i=1}^{N} (y_i - \eta(t_i, x_i))^T R_i^{-1} (y_i - \eta(t_i, x_i)). \tag{1.9}$$

The statistical background of the LS method, including its connection with maximum likelihood estimation, will be presented in Chapter 3. Efficient numerical methods for this kind of estimation problems have become mature techniques. In this connection, the

Gauss-Newton method combined with multiple shooting has proved to be efficient (see Section 1.8). Despite advances, many challenges still prevail like problems of massive data, real-time requirements, local minima. Besides that, quantifying the accuracy of the estimates is of primary importance. This can be approached by computing their covariance matrices and confidence regions. These involve calculating the derivatives of solutions of ODEs with respect to initial states and parameters, which can be expensive. Fast numerical methods, especially in the context of online estimation, need to be developed to deal with those challenges.

1.3 Gauss-Newton method for parameter estimation

The constrained Gauss-Newton method, or Gauss-Newton (GN) method for short, is a powerful and perhaps the most popular method for solving parameter estimation problems numerically. It has been studied extensively in the works of Bock [11] and Schlöder [78]. One of the favorable properties of the GN method argued by the authors is that, upon convergence, the solution is stable under variations of noise. Moreover, the GN method makes up the fundamentals of nonlinear OED. Let us here sketch the basics of the GN method. Without loss of generality, assume in (1.8) that for $i = 1, 2, ..., N$, $R_i = \mathbb{I}_{n_y}$. Set

$$F_1(v) = (y_i - \eta(t_i, x_i))_{i=1}^N, \quad v = (x_0, p)$$

The parameter estimation problem can be rewritten in the form

$$\begin{aligned} \min_{v \in \mathbb{R}^{n_v}} \ & \frac{1}{2}\|F_1(v)\|_2^2, \\ \text{s.t. } & F_2(v) = 0, \end{aligned} \tag{1.10}$$

where $F_1(v) \in \mathbb{R}^N$, $F_2(v) \in \mathbb{R}^{N_2}$ ($N_2 \leq n_v$). For simplicity, we consider constraints only in the form of equalities, expressed by F_2. In the case of inequality constraints, one can use an active set method and locally treat the problem as a problem with only equality constraints. Henceforth we denote by J_1, J_2 the Jacobian matrix of F_1, F_2, respectively.

GN algorithm: Starting with an estimate v_0, for $k = 0, 1, 2, ...$

1. Solve the linearized problem

$$\begin{aligned} \min_{\Delta v_k} & \frac{1}{2}\|J_1(v_k)\Delta v_k + F_1(v_k)\|_2^2, \\ & J_2(v_k)\Delta v_k + F_2(v_k) = 0. \end{aligned} \tag{1.11}$$

2. Compute the new iterate

$$v_{k+1} = v_k + \Delta v_k.$$

3. If v_{k+1} satisfies some termination criterion, stop the algorithm and the solution is $\hat{v} = v_{k+1}$. If not, continue with 1.

A possible termination criterion is $\|\Delta v_k\| \leq \texttt{tol}$ with a tolerance $\texttt{tol} > 0$ (Schlöder [78]). Suppose that the CQ (constraint qualification) and PD (positive definiteness) hold, i.e.,

- (CQ). Rank $J_2(v) = N_2$.

- (PD). Rank $\begin{pmatrix} J_1(v) \\ J_2(v) \end{pmatrix} = n_v$

for all $v \in \mathbb{R}^{n_v}$. Then the solution of (1.11) admits an explicit form

$$\begin{aligned} \Delta v_k &= -\begin{pmatrix} \mathbb{I}_N & 0 \end{pmatrix} \begin{pmatrix} J_1(v_k)^T J_1(v_k) & J_2^T(v_k) \\ J_2(v_k) & 0 \end{pmatrix}^{-1} \begin{pmatrix} J_1^T(v_k) & 0 \\ 0 & \mathbb{I}_{N_2} \end{pmatrix} \begin{pmatrix} F_1(v_k) \\ F_2(v_k) \end{pmatrix} \\ &= -J^+(v_k) \begin{pmatrix} F_1(v_k) \\ F_2(v_k) \end{pmatrix} \end{aligned}$$

where

$$J^+(v_k) = \begin{pmatrix} \mathbb{I}_N & 0 \end{pmatrix} \begin{pmatrix} J_1^T(v_k) J_1(v_k) & J_2^T(v_k) \\ J_2(v_k) & 0 \end{pmatrix}^{-1} \begin{pmatrix} J_1^T(v_k) & 0 \\ 0 & \mathbb{I}_{N_2} \end{pmatrix}.$$

The matrix $J^+(v_k)$ is called a generalized inverse of the matrix $\begin{pmatrix} J_1(v_k) \\ J_2(v_k) \end{pmatrix}$. A linear approximation of the covariance matrix of the solution v^* of (1.10) can be computed by

$$C = \begin{pmatrix} \mathbb{I}_N & 0 \end{pmatrix} \begin{pmatrix} J_1^T J_1 & J_2^T \\ J_2 & 0 \end{pmatrix}^{-1} \begin{pmatrix} J_1^T J_1 & 0 \\ 0 & \mathbb{I}_{N_2} \end{pmatrix} \begin{pmatrix} J_1^T J_1 & J_2^T \\ J_2 & 0 \end{pmatrix}^{-T} \begin{pmatrix} \mathbb{I}_N \\ 0 \end{pmatrix}.$$

where F_1, F_2, J_1, J_2 are evaluated at v^*. In case there is no constraint, C reduces to $C = (J_1^T J_1)^{-1}$.

The covariance matrix plays an important role in assessing the quality of parameter estimates. To make it clear, we consider the concept of *confidence region* (see Rice [73]). Take $0 \leq \alpha \leq 1$. Regarding the problem of estimating the parameters $v \in \mathbb{R}^{n_v}$, a $(100\alpha)\%$ confidence region is a random region in $\mathbb{R}^{n_v}$ depending on measurements with the following property: If we form many regions based on random measurements, then on average $(100\alpha)\%$ of these regions will contain the true parameters. It is shown in Bock [11] and Körkel [46] that a linear approximation of a $(100\alpha)\%$ confidence region, denoted by $G_L(\alpha, \hat{v})$, satisfies

$$G_L(\alpha, \hat{v}) \subseteq [\hat{v}_1 - \theta_1, \hat{v}_1 + \theta_1] \times ... \times [\hat{v}_{n_v} - \theta_{n_v}, \hat{v}_{n_v} + \theta_{n_v}].$$

where $\hat{v}$ is the solution provided by the GN algorithm and $\theta_i = \gamma(\alpha)\sqrt{C_{ii}}$. Here $\gamma(\cdot)$ is the χ^2-distribution with $(n_v - N_2)$ degrees of freedom and C is evaluated at $\hat{v}$. Hence, the covariance matrix is a statistical quantity allowing us to assess the accuracy of the estimates. OED aims to reduce this covariance matrix in a suitable sense, see Chapter 4.

In reality, especially in the context of NMPC, measurements are collected in a long period and often are increasing in number over time. A natural question arises: what are the asymptotic properties of the estimates? There are results on this topic which claim that the estimates are consistent and approach a normal distribution, see Chapter 3. Sufficient conditions for such results sometimes do not hold. It is therefore of interest to investigate the asymptotic properties in this case. On the other hand, as the sample sizes grow, the estimation problem may become ill-posed. This leads to numerical difficulties which need to be tackled. These issues are addressed in Chapter 3.

1.4 Optimal Experimental Design

By utilizing the description of noise, we can investigate statistical properties of the estimates of parameters such as covariances, confidence regions. A small confidence region

is an indicator of good quality and reliability of estimates. It is desirable to obtain estimates with confidence regions as small as possible. The delicacy here is that, without knowing realizations of measurements and solving the estimation problem beforehand, we can still approximate the statistical properties of estimates. In view of the Gauss-Newton method for parameter estimation, the covariance matrix of the estimates does not depend on concrete values of measurements, but depends on controls and time points to measure. By skillfully choosing controls and measuring times, which are called experimental conditions, we can drastically reduce the covariance of the estimates and therefore obtain small confidence regions. The reliability of the estimates is improved. OED aims to minimize some scalar function $\mathscr{K}(C)$ acting on the set of possible covariance matrices. Several well-known ones are

$$\mathscr{K}_D(C) = (\det(C))^{\frac{1}{n_v}} \quad \text{(D-criterion)},$$

$$\mathscr{K}_A(C) = \frac{1}{n_v}\text{Trace}(C) \quad \text{(A-criterion)},$$

$$\mathscr{K}_M(C) = \max\{\sqrt{C_{ii}}, \quad i = 1, 2, ..., n_v\} \quad \text{(M-criterion)}.$$

In this thesis, we are also interested in the *G-criterion*,

$$\mathscr{K}_G(C) = g^T C g \quad \text{(G-criterion)}$$

for some $g \in \mathbb{R}^{n_v}$. The idea is that we pay attention to particular directions or particular combinations of the parameters. It will be seen in Chapter 6 that in role of g, the sensitivity of the optimal nominal objective value of OCPs with respect to unknown parameters and initial states is useful for Dual Control.

OED for linear models has been studied thoroughly in the literature, see for example Pukelsheim [69]. In spite of that, there are many difficulties in the nonlinear case. For the first one, the computational effort is demanding. The covariance matrix involves the derivatives of the measurement function with respect to parameters. In models based on nonlinear differential equations, it can be hard and expensive to compute those derivatives. Körkel [46] presents advanced numerical methods for nonlinear OED together with a sophisticated software implementation, which has been employed successfully in industrial applications. Another essential difficulty of nonlinear OED is that the covariance matrix depends on specific values of the parameters. The results obtained by solving an OED problem is optimal for the current estimates, not for the true parameters. Some strategies to overcome this difficulty are studied, for example, sequential OED (Körkel et al. [47]), robust OED (Körkel [46], Körkel et al. [48]). We note that it is possible to formulate the problem of sequential OED as a stochastic OCP and handle it with Dual Control techniques, see Chapter 5. Moreover in Chapter 4, we present several interesting results on finite support designs in order to exhibit the connection between continuous designs and sampling designs.

1.5 Nonlinear Model Predictive Control and Moving Horizon Estimation

Nonlinear Model Predictive Control (NMPC) is a popular strategy to implement feedback control for nonlinear systems. Often there are unknown parameters and disturbances in the process to be controlled. We want to know the parameters and the states

at the current time on the one hand, and to compute control actions in a feedback form on the other hand. For this purpose, we measure the states or a part of the states of the system, called outputs, at discrete time points, called sampling instants. Using those measurements, we estimate the parameters and the states at the current time and use them for prediction to compute control actions. See Chapter 7 for a concrete NMPC setup. The NMPC procedure can be described along Figure 1.1 as follows.

At time t_n, the estimates of the states x_n, and of the unknown parameters p are available. Using the model, we can predict the states of the process on the time horizon $[t_n, t_{n+N}]$. We solve an OCP by discretizing the controls u, e.g., piecewise constant, on a uniform grid $t_n < t_{n+1} < ... < t_{n+N}$ and receive an optimal control sequence $(u_0^*, u_1^*, ..., u_{N-1}^*)$. Only the first few elements of this control sequence are applied to the system, say, $(u_0^*, u_1^*, ..., u_{N_c-1}^*)$ $(1 \leq N_c < N)$. Depending on the arrangement of sensors, the outputs at some time points in $(t_n, t_{n+N_c}]$ are obtained, denoted by $y_{n_1}, ..., y_{n_m}$. Using the latest N_e measurements, we estimate the states x_{n+N_c} at t_{n+N_c} as well as the parameters. We then repeat the procedure with the next horizon $[t_{n+N_c}, t_{n+N_c+N}]$. Figure 1.1 illustrates a special NMPC scheme where measurement times are exactly on the grid.

In this work, we call N, N_c, N_e the *prediction horizon*, *control horizon* and *estimation horizon*, respectively. Since the time intervals for estimation are shifted at each step and because we require the number of measurements not to exceed N_e, this estimation strategy is called *moving horizon estimation* (MHE). Likewise the process is controlled on moving horizons. NMPC is therefore called *moving horizon control* or *receding horizon control*. There is also an NMPC strategy on a fixed interval called batch NMPC where the horizons are contracted or shrunk. We will apply a batch NMPC for time optimal control in Chapter 8.

NMPC is not feedback control in the original meaning, however. At time t_n, for instance, the control action u_0^* can be considered as feedback to x_n but then in $(t_n, t_{n+1}]$, ..., (t_{n+N_c-1}, t_{n+N_c}) the control is of open loop. This is an intrinsic property of the computer-controlled systems due to their digital nature. On the other hand, the delay caused by computational time of solving OCP and MHE may be significant. For successful implementation of NMPC in practice, it is decisive to develop fast solvers or to handle delays. Among the most striking results in this direction are the *real-time iterations* (Diehl [19]), *advanced-step NMPC* (Bock et al. [12], Zavala and Biegler [87]), *multilevel iterations* (Albersmeyer et al. [2], Kirches et al. [45]). See Section 1.9 for more details.

1.6 Control, estimation and the separation principle

At each NMPC step, the states and parameters have to be estimated. An OCP is solved based on these estimates. Since estimation problems and OCPs follow one after another and since the estimates are not exact, the following questions arise naturally:

- Should the OCP take into account the uncertainties in the estimates resulting from the previous estimation problem or just consider those estimates as true values?

- From the viewpoint of OED, the control has an impact on the uncertainties of the estimates. Should the controls attempt to improve the quality of the estimates while minimizing the objective function specified by the OCP?

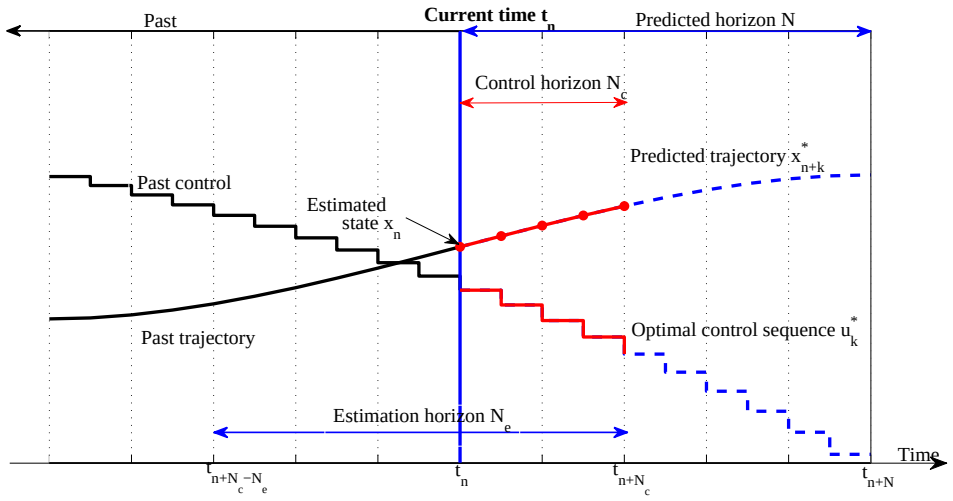

Figure 1.1: An NMPC scheme on receding horizons. At the current time t_n, with the estimate of the state x_n, an OCP is solved on the horizon $[t_n, t_{n+N}]$ in order to obtain a control sequence $(u_0^*, u_1^*, ..., u_{N-1}^*)$. The elements $u_0^*, ..., u_{N_c-1}^*$ are applied to the system. The outputs at $t_{n+1}, ..., t_{n+N_c}$ are measured and the states at t_{n+N_c} are estimated based on the latest N_e measurements on $[t_{n+N_c-N_e}, t_{n+N_c}]$. The procedure is then repeated on the shifted horizon $[t_{n+N_c}, t_{n+N_c+N}]$

For the first question, if we consider the estimates, which we call *nominal values*, as the true values, we thus implicitly make use of the *Certainty Equivalence* principle (see the next section). This strategy is regarded as *nominal NMPC*. On the other hand, if we treat estimates as random variables, we are to use methods from stochastic optimization. A solution to this problem is often too complicated and too expensive to be obtained in closed form. Some types of approximation that utilize the covariance matrix of the estimates need to be employed.

For the second question, the controller undertakes two tasks: *performance control* and *information gain*. The *separation principle* refers to the independence of the two tasks. This means that the optimal controller can be decomposed into two separate parts: An estimator, which gives the estimates of states and parameters; and an actuator, which computes control inputs based on those estimates, see Bertsekas [10]. The separation principle is valid for a limited number of cases such as Linear Quadratic Gaussian (LQG). In general, especially for nonlinear systems, it does not hold. To some degree, optimality of performance control is lost in the effort of estimating. In return, if we soon get good estimates of states and parameters, the control actions in the future is likely to be improved, leading to a better overall performance.

Nominal NMPC has proved its simplicity and effectiveness in practice. This is in part because the performance control task and the information gain task often go in the same direction. However, this is not always the case. There are situations in which the two tasks are conflicting. Nominal NMPC can result in infeasibility and nonrobustness in practical applications. We will illustrate these issues by concrete examples in Chapter 8. Success in justifying their relationship and making a suitable balance are important for improving the performance of NMPC. This is the main idea of Dual Control that will be clarified in Section 1.10 and Chapters 5–6.

1.7 Certainty Equivalence principle

The *Certainty Equivalence* principle stands for the interchangeability of taking minimization and taking expectation.

For convenience, we recall the concept of conditional expectations, see for example, Durrett [22]. Suppose that $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ are two random vectors (RVs) defined on the probability space $(\Omega, \mathbb{P})$, where $\mathbb{P}$ is the corresponding probability measure. Furthermore, let $f_{XY}(x, y)$ be their joint probability density function (pdf) and $f_X(x), f_Y(y)$ their marginal pdfs. The conditional expectation of X given $Y = y$ is denoted by $\mathbb{E}\big[X|Y = y\big]$ and has the pdf given by $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$ if $f_Y(y) \neq 0$ and 0 otherwise. If $\Phi(\cdot)$ is a function of X, then the conditional expectation of $\Phi(X)$ given $Y = y$ can be computed as follows

$$\mathbb{E}\left[\Phi(X) \;\middle|\; Y = y\right] = \int_{\mathbb{R}^n} \Phi(x) f_{X|Y}(x|y) dx.$$

Obviously, $h(y) = \mathbb{E}\left[\Phi(X) \;\middle|\; Y = y\right]$ induces a RV $h(Y)$, denoted by $\mathbb{E}\left[\Phi(X) \;\middle|\; Y\right]$. For its expectation, the following identities hold (see e.g., Durrett [22])

$$\begin{aligned}
\mathbb{E}(h(Y)) &= \int_{\mathbb{R}^m} h(y) f_Y(y) dy = \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} \Phi(x) f_{X|Y}(x|y) f_Y(y) dx dy \\
&= \int_{\mathbb{R}^n} \Phi(x) \left(\int_{\mathbb{R}^m} f_{X|Y}(x|y) f_Y(y) dy \right) dx \\
&= \int_{\mathbb{R}^n} \Phi(x) \left(\int_{\mathbb{R}^m} f_{XY}(x, y) dy \right) dx = \int_{\mathbb{R}^n} \Phi(x) f_X(x) dx \\
&= \mathbb{E}(\Phi(X)).
\end{aligned}$$

In addition to the random vector X, suppose that $\varphi : \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}$ is a real valued function, where $\mathbb{U}$ is subset of $\mathbb{R}^{n_u}$. The question of interest is whether the following two problems

$$\min_u \mathbb{E}\varphi(X, u) \quad \text{and} \quad \mathbb{E} \min_u \varphi(X, u)$$

are equivalent. We would like to have controls u in feedback form given by the *control policy* $\mu = \mu(X)$ that is a function mapping $\mathbb{R}^n$ into $\mathbb{U}$. Under certain conditions, the Certainty Equivalence principle holds, see Åström [4, Chapter 8]. In fact, suppose that for each $z \in \mathbb{R}^n$, there exists a $u^* = \mu^*(z) \in \mathbb{U}$, not necessarily unique, such that

$$\varphi(z, \mu^*(z)) = \min_{u \in \mathbb{U}} \varphi(z, u).$$

Then for any control policy $\mu(X)$, and any $\omega \in \Omega$ we have

$$\varphi(X(\omega), \mu(X(\omega))) \geq \varphi(X(\omega), \mu^*(X(\omega))).$$

It follows that

$$\mathbb{E}\varphi(X, \mu(X)) \geq \mathbb{E}\varphi(X, \mu^*(X)).$$

This inequality holds for any $\mu(X)$, hence

$$\min_{\mu(X)} \mathbb{E}\varphi(X, \mu(X)) \geq \mathbb{E}\varphi(X, \mu^*(X)).$$

On the other hand, since $\mu^*(X)$ is a control policy, the reverse inequality is also valid. Thus we have proved

$$\min_{\mu(X)} \mathbb{E}\varphi(X, \mu(X)) = \mathbb{E}\varphi(X, \mu^*(X)) = \mathbb{E} \min_{\mu(X)} \varphi(X, \mu(X)). \tag{1.12}$$

A common case in practice is when the states X are not completely accessible. Instead only measurements Y which are an n_y dimensional random vector are available. The minimization problem is formulated as $\min_{\mu(Y)} \mathbb{E}\gamma(X, Y, u)$ where $\gamma : \mathbb{R}^n \times \mathbb{R}^{n_y} \times \mathbb{U} \to \mathbb{R}$ is a function to be minimized and the expectation is evaluated with respect to all random vectors involved. Control policies are functions that map $\mathbb{R}^{n_y}$ into $\mathbb{U}$, i.e., $u = \mu(Y)$. For each $y \in \mathbb{R}^{n_y}$, $u \in \mathbb{U}$, consider the conditional expectation $\ell(y, u) = \mathbb{E}\left[\gamma(X, Y, u) \middle| Y = y\right]$ and suppose that there exists $\mu^*(y) = \operatorname{argmin}_{w \in \mathbb{U}} \ell(y, w)$. Similar arguments as in the proof of (1.12) show that

$$\min_{\mu(Y)} \mathbb{E}\gamma(X, Y, \mu(Y)) = \mathbb{E}\ell(Y, \mu^*(Y)) = \mathbb{E} \min_{\mu(Y)} \mathbb{E}\left[\gamma(X, Y, \mu(Y)) \;\middle|\; Y\right]. \tag{1.13}$$

As a special case, let $\varphi(z, u)$ be a quadratic function of z and u with a symmetric positive definite Hessian matrix R with respect to u,

$$\varphi(z, u) = u^T R u + 2u^T K z + z^T Q z, \quad R \succ 0.$$

Since $\mu^*(z) = -R^{-1}Kz$ uniquely minimizes φ with respect to u, the Certainty Equivalence principle (1.12) is valid for this case. This is the background for the *Linear Quadratic Gaussian* (LQG), a fundamental problem in stochastic control theory. In this case, both the separation principle and the certainty equivalence principle hold (see Åström [4], Bertsekas [10]). At each step, the Kalman filter yields the best estimate of states. The *Linear Quadratic Regulator* (LQR) uses this estimate to compute optimal control. We will present an accessible derivation of LQR in Chapter 5.

1.8 Direct Multiple Shooting method

Multiple Shooting (MS) has originated from efforts to solve boundary value problems (see Osborne [65]). Instead of solving a system of ODEs on the whole interval and matching boundary conditions at the starting point and the end point like single shooting, MS introduces a grid consisting of intermediate points between the starting point and the end point and impose equality constraints to enforce the continuity of the trajectories (see Stoer and Burlisch [82], Bock and Plitt [13]). Nowadays, multiple shooting refers to direct MS applied to solve OCPs. Regarding OCP (1.1)–(1.5) with t_f fixed, we consider the grid $t_0 < t_1 < t_2 ... < t_N = t_f$. Suppose that the controls are piecewise constant, i.e.,

$$u(t) = u_i \in \mathbb{R}^{n_u}, \quad t \in [t_i, t_{i+1}).$$

We introduce new variables $s_i \in \mathbb{R}^{n_x}$ that represent the values of the states at t_i. OCP (1.1)–(1.5) can be formulated in discretized form as follows

$$\min_{u_i, s_i} \Phi(s_N) + \sum_{i=0}^{N-1} L_i(s_i, u_i) \tag{1.14}$$

subject to

$$s_{i+1} = x(t_{i+1}; t_i, s_i, u_i, p), \quad i = 0, 1, ..., N-1, \tag{1.15}$$

$$s_0 = x_0, \tag{1.16}$$

$$u_i \in \mathbb{U}, \quad i = 0, 1, ..., N-1. \tag{1.17}$$

$$\Psi(t_N, s_N) = 0, \tag{1.18}$$

Here as in Section 1.1, the notation $x(t; t_i, s_i, u_i, p)$ stands for the solution of the IVP (1.2) with $x(t_i) = s_i$ and $u(t) \equiv u_i$ for $t \in [t_i, t_{i+1})$. Furthermore we set

$$L_i(s_i, u_i) = \int_{t_i}^{t_{i+1}} L(t, x(t), u(t))dt$$

with $x(t) = x(t; t_i, s_i, u_i, p)$ and $u(t) \equiv u_i$ for $t \in [t_i, t_{i+1})$. Constraints (1.15) are called *matching conditions*. See Figure 1.2 for a visualization of a direct multiple shooting scheme. We remark that path constraints in MS require special treatments that are presented in Potschka [67].

We have introduced additional variables s_i together with corresponding matching constraints. This results in a large but well structured nonlinear program. Efficient methods for dealing with this type of large nonlinear program such as condensing techniques, internal numerical differentiation have been developed and gained remarkable success. The advantages of direct MS over single shooting are therefore tremendous, see Bock and Plitt [13], Schlöder [78]. They include the reduction of nonlinearity (because ODEs are solved in much shorter intervals), broader ranges of local convergence. MS is also applicable to parameter estimation, see Bock [11], Schlöder [78]. In this case it has an additional benefit by making use of the measurements for initializing intermediate values of the optimal solution.

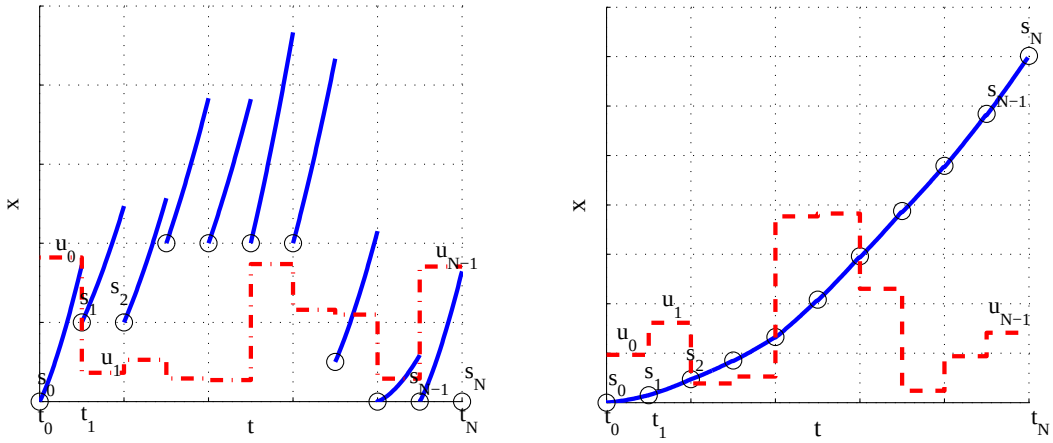

Figure 1.2: This figure illustrates a direct multiple shooting scheme. At initialization (left), the trajectory is only piecewise continuous due to the mismatch of the solution of the ODEs on an interval and the next interval. After fulfilling the matching conditions, the trajectory becomes continuous (right).

1.9 Real-Time NMPC

In conventional NMPC schemes, we assume that OCPs and Moving Horizon Estimation (MHE) problems can be solved instantly and that the computed control sequence and

the estimated states and parameters are applied to the process immediately. In practice, however, there are always delays. This can be elucidated as follows. At time t_n, the initial states that we have obtained by carrying out MHE may not be the estimated states at t_n, rather at some time point earlier, namely $t_n - \delta_{\text{MHE}}$. We then solve an OCP which takes an amount of time δ_{OCP} to complete. The resulting optimal control sequence is applied to the system at time $t_n + \delta_{\text{OCP}}$ with the initial states $x(t_n + \delta_{\text{OCP}})$ instead of t_n and x_n. The more computational time spent on solving MHE and OCP, the longer is the delay. This delay certainly makes the optimal control sequence no longer optimal, i.e., sub-optimal, and even worse, can result in infeasibility and instability. In order to fully explore the advantages of NMPC, strategies to reduce delays or handle delays must be studied and implemented. This initiates the concept of real-time iterations or real-time NMPC which has received intensive research in recent years, Diehl [19], Diehl et al. [21].

Often a few iterations can yield good approximation of the optimal solution. Hence instead of iterating until convergence, it is reasonable to let the solver carry out only a few iterations which satisfy certain properties such as reduction of the objective function. Also between consecutive OCPs, there is a close relationship. Not only their structure and their data but also their solutions are likely to be similar. There is a good chance to make use of the computed solution of an OCP to initialize the next OCP. This could immediately ensure feasibility and good approximation of its optimal solution. The initialization technique includes

- Warm-start strategy: The whole computed control sequence $u_0, u_1, ..., u_{N-1}$ becomes the initialization for the next OCP.
- Shift strategy: The next OCP is initialized by $(u_1, ..., u_{N-1}, u_{N-1})$.

On the other hand, for an OCP, only the initial state x_0 is unknown. An approximated solution of an OCP can be obtained in two phases: *preparation phase* and *feedback phase*. In the preparation phase, by using the *initial value embedding* strategy, we treat x_0 as a parameter and add a linear constraint $s_0 - x_0 = 0$ to the constraints of the OCP. The computation of linear constraints and approximation of quantities based on previous states can be done prior to the knowledge of x_0. In the feedback phase, when x_0 is known, we substitute x_0 into the computed quantities and can obtain a new iteration instantly. In a recently developed procedure called *multilevel iterations*, one can choose several options at a particular OCP like retaining, updating in part or completely updating the data of the previous OCP. See Bock et al. [12], Diehl [19], Albersmeyer et al. [2], Kirches et al. [45] and Frasch et al. [30] for details. It is also possible to choose the iteration types adaptively, based on some measures on the deviation of the current OCP and the next OCP.

1.10 Dual Control for NMPC

NMPC combined with Moving Horizon Estimation (MHE) is one of the promising ways to deal with uncertainties and hence to control uncertain systems. The estimation procedure helps to improve knowledge about the system and converge to the true parameters. If the estimates become more and more accurate, better control actions in the future can be obtained. At the same time, OED has proven to be useful and effective for enhancing the accuracy of the estimates before measurements are actually taken. OED is also of vital importance in the face of a limited number of costly measurements, which

is often the case in practice. There are two aspects that the controls should take into account. The first one is to optimize the original objective function while satisfying the constraints. We call it the *performance control task*. The second one is to get information about the process in order to get better estimates when new measurements are available. We call it the *information gain task*. By gaining more information and obtaining good estimates of parameters and states on time, control actions tend to be better in the future. This leads to an improvement of the overall performance. The two tasks are sometimes conflicting, sometimes they support each other. In general, the interplay between them is complicated. It is beneficial to understand their relationship and strike a balance. This is the key idea of *Dual Control*. In the context of NMPC, at each step, the original OCP should be modified to incorporate future information. In fact, the problem of Dual Control was considered by Feldbaum [26] in the 1960s in the context of signal processing. The stochastic equations for the optimal solutions based on nested conditional expectations exist, Feldbaum [26], Åström [4], Bertsekas [10], but are too intricate to be solved efficiently. In fact, the classical approaches to Dual Control have faced hindrance in practical applications, presumably because of the *curse of dimensionality* and difficulties in computing conditional expectations. This calls for approximation together with advanced numerical methods to cope with complexity. Chapter 5 will give a comprehensive introduction to Dual Control together with a survey on recent developments. In Chapter 6 we will present new approaches to Dual Control based on OED together with the statistical background.

Chapter 2

Sensitivity Analysis of Optimal Control Problems

2.1 Introduction

Two milestones of the optimal control theory are the Pontryagin Minimum Principle (PMP) and the Hamilton-Jacobi-Bellman Equation (HJBE). Both are used to characterize optimal solutions of Optimal Control Problems (OCPs). While the PMP considers how variations of the controls affect the objective function, the HJBE keeps track of the behavior of the optimal objective value as the problem parameters such as initial states, initial time, constant parameters vary. Those different views lead to significantly distinguishing features which can be summarized as follows.

- PMP provides necessary conditions. The optimal control is obtained as a function of time.
- HJBE provides sufficient conditions. The value function is assumed to be differentiable. The optimal control is given in feedback form.

We introduce in this chapter a rather general form of the PMP and the HJBE which fits our framework. Based on that, extensions and generalizations can be derived with minor efforts. After stating the main theorems, we present several illustrative examples that provide analytic solutions to the numerical examples in Chapter 8. We then carry out a sensitivity analysis of OCPs, mainly with respect to initial states and constant parameters. This analysis is useful for understanding the effect of uncertainties on OCPs, suggesting methods to improve control performance. We consider OCP (1.1)–(1.5), which we recall here for ease of reference.

$$\min_{u(\cdot),x(\cdot),t_f} J = \Phi(x(t_f)) + \int_{t_0}^{t_f} L(t, x(t), u(t))dt \tag{2.1}$$

$$\dot{x}(t) = f(t, x(t), u(t)), \quad t \in [t_0, t_f], \tag{2.2}$$
$$x(t_0) = x_0; \tag{2.3}$$
$$u(t) \in \mathbb{U} \subseteq \mathbb{R}^{n_u}, \quad t \in [t_0, t_f]; \tag{2.4}$$
$$\Psi(t_f, x(t_f)) = 0 \in \mathbb{R}^{n_\Psi}. \tag{2.5}$$

Note that we fixed initial time t_0 and allow final time t_f to vary. By a feasible final time, we understand a time t_f such that there exists a *piecewise continuous* function $u : [t_0, t_f] \to \mathbb{U}$ for which conditions (2.2)–(2.5) are fulfilled. In this case, $u(\cdot)$ is also called a feasible control without explicitly mentioning t_f. Note that we have omitted parameters p for brevity, assuming that they are given.

2.2 Pontryagin Minimum Principle and the necessary conditions

The Pontryagin Minimum Principle (PMP) (also called maximum principle) was conjectured by Pontryagin around 1955. The rigorous proof of the PMP was given by Boltjanskii in 1958 (see Zeidler [88]). We assume that all functions L, f, Φ, Ψ have continuous first-order partial derivatives with respect to all arguments. Define the Hamiltonian of the system as a function $H : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R} \to \mathbb{R}$ with

$$H(t, x, u, \lambda, \lambda_0) = \lambda_0 L(t, x, u) + \lambda^T f(t, x, u).$$

Theorem 2.1. (Zeidler [88, p. 424]) *If $u^*(t), x^*(t), t_f$ solve OCP (2.1)–(2.5), then there exist a constant $\lambda_0 \geq 0$, a constant vector $\nu \in \mathbb{R}^{n_\Psi}$ and a continuously differentiable vector function $\lambda(t) \in \mathbb{R}^n$, not all zero, such that*

$$H(t, x^*(t), u^*(t), \lambda(t), \lambda_0) = \min_{w \in \mathbb{U}} H(t, x^*(t), w, \lambda(t), \lambda_0) \tag{2.6}$$

for $t_0 \leq t \leq t_f$ at which u^* is continuous, where $\lambda(t)$ satisfies the adjoint equation

$$\dot{\lambda}(t) = -H_x^T(t, x^*(t), u^*(t), \lambda(t), \lambda_0) \tag{2.7}$$

with the end time conditions

$$\lambda(t_f) = \lambda_0 \Phi_x^T(x^*(t_f)) + \Psi_x^T(t_f, x^*(t_f))\nu. \tag{2.8}$$

Furthermore, there exists a differentiable function $p_0(t) : [t_0, t_f] \to \mathbb{R}$ such that at all continuity points t of $u^*(t)$, it holds that

$$\begin{aligned} p_0(t) &= -H(t, x^*(t), u^*(t), \lambda(t), \lambda_0), \\ \dot{p}_0(t) &= -H_t(t, x^*(t), u^*(t), \lambda(t), \lambda_0), \quad p_0(t_f) = \Psi_t^T(t_f, x(t_f))\nu. \end{aligned} \tag{2.9}$$

Remark 2.1. *The degenerate case occurs when $\lambda_0 = 0$. In this case, the optimality conditions are independent of the integral to be minimized. On the contrary, if $\lambda_0 > 0$, by setting $\tilde{L}(\cdot) = \lambda_0 L(\cdot)$, one can choose $\lambda_0 = 1$, without loss of generality. Then the optimal solutions will depend on the objective function, which is physically more reasonable. In applying the PMP, it is wise to try $\lambda_0 = 1$ first. Hence, from now on, we always take $\lambda_0 = 1$ and remove it from the Hamiltonian.*

If there are no constraints on the control, i.e., $\mathbb{U} = \mathbb{R}^{n_u}$, (2.6) is equivalent to

$$H_u(t, x^*(t), u^*(t), \lambda(t)) = 0$$

or equivalently

$$L_u(t, x^*(t), u^*(t)) + \lambda^T(t) f_u(t, x^*(t), u^*(t)) = 0. \tag{2.10}$$

This is also valid if the minimum of H with respect to u lies in the interior of $\mathbb{U}$.

Remark 2.2. *The PMP supplies a set of necessary conditions, which an optimal control if existing has to satisfy. It may happen that there are several $u(\cdot)$ fulfilling the PMP, but some of them are solutions while the others are not or even none of them, simply because the OCP has no solutions. A strategy to solve an OCP is to use the PMP to determine candidates of solutions and then prove the existence of solutions or directly show that a candidate is actually a solution. Example* 2.1 *provides a detailed illustration. We note that sufficient conditions for solutions of OCPs will be presented later in Section* 2.3.

Example 2.1. *Consider the following OCP*

$$\min J(u) = \int_0^{t_f} \left(\frac{u^2}{2} - x\right) dt \tag{2.11}$$

subject to

$$\dot{x}(t) = u, \quad x(0) = 0.$$

The Hamiltonian of the system is

$$H(t, x, u, \lambda) = \frac{u^2}{2} - x + \lambda u$$

Since there is no constraint on u, H is minimized with respect to u if and only if $H_u(t, x, u, \lambda) = u + \lambda = 0$. If follows that $u = -\lambda$. Furthermore, the adjoint equation has the form $\dot{\lambda}(t) = -H_x = 1$. Therefore

$$\lambda(t) = t + a, \quad u(t) = -t - a$$

where $a = \lambda(0)$. We easily get $x(t) = -\dfrac{t^2}{2} - at$. In addition, (2.8) *yields $\lambda(t_f) = t_f + a = 0$. Hence $a = -t_f$.*

Consider now the case when the final time t_f is free. It follows from (2.9) *that*

$$\frac{(t_f + a)^2}{2} + \left(\frac{t_f^2}{2} + at_f\right) - (t_f + a)^2 = 0.$$

We obtain $a^2 = 2$ and then $a = \pm\sqrt{2}$. Because $-a = t_f \geq 0$, the only set of functions satisfying the PMP corresponds to $a = -\sqrt{2}$. OCP (2.11) *does not possess any solution, however. In fact, with the control $\tilde{u} \equiv b > 0$, $x(t) = bt$ and the objective function is*

$$J(\tilde{u}) = \int_0^{t_f} \left(\frac{b^2}{2} - bt\right) = \frac{b^2 t_f}{2} - \frac{bt_f^2}{2}$$

which tends to $-\infty$ as $t_f \to \infty$.

On the other hand, suppose that t_f is fixed and positive, for example $t_f = 1$. Thus $a = -1$. The control $u^*(t) = -t + 1$ fulfills the PMP. If we are able to prove the existence of a solution, we can conclude that $u^*(t)$ is the unique optimal control. This can be done straightforwardly. Indeed, let $u_h(t) = u^*(t) + h(t)$ be an arbitrary feasible control. Writing $x(t) = \int_0^t u(\tau)d\tau$, we obtain

$$\begin{aligned} J(u_h) - J(u^*) &= \int_0^1 \left(\frac{(u^*(t) + h(t))^2 - u^*(t)}{2} - \int_0^t \big(u^*(\tau) + h(\tau) - u^*(\tau)\big)d\tau\right) dt \\ &= \int_0^1 \frac{h^2(t)}{2} dt + \int_0^1 u^*(t)h(t)dt - \int_0^1 \int_0^t h(\tau)d\tau dt. \end{aligned}$$

Applying integration by parts, we have

$$\begin{aligned}\int_0^1 u^*(t)h(t)dt &= \int_0^1 (1-t)h(t)dt = \int_0^1 (1-t)d\left(\int_0^t h(\tau)d\tau\right)\\ &=(1-t)\int_0^t h(\tau)d\tau\Big|_0^1 - \int_0^1\int_0^t h(\tau)d\tau d(1-t)\\ &=\int_0^1\int_0^t h(\tau)d\tau dt.\end{aligned}$$

Therefore

$$J(u_h) - J(u^*) = \int_0^1 \frac{h^2(t)}{2}dt.$$

Hence $J(u_h) \geq J(u^*)$ *and* $J(u_h) = J(u^*)$ *if and only if* $h(t) \equiv 0$ *or* $u_h(t) \equiv u^*(t)$. *As a result,* $u^*(t)$ *is the unique solution of the OCP at hand.*

The next two examples are of practical interest for which detailed explanations of physical quantities will be given in Chapter 8.

Example 2.2. *Consider a modified version of the rocket car problem*

$$\min_{u(\cdot),x(\cdot),t_f} t_f$$

subject to

$$\begin{cases}\dot{x}_1(t) = x_2(t),\\ \dot{x}_2(t) = au_1(t) - bu_2(t),\\ x_1(0) = 0, x_2(0) = 0,\end{cases} \tag{2.12}$$

where t_f *is a terminal time such that*

$$x_1(t_f) = 1,\ x_2(t_f) = 0.$$

The controls are subject to constraints

$$0 \leq u_1(t), u_2(t) \leq 2 \text{ for all } 0 \leq t \leq t_f.$$

Here a *and* b *are positive parameters. The objective function can be written as*

$$J = \int_0^{t_f} 1dt.$$

Thus the Hamiltonian takes the form

$$H(t, x, u, \lambda) = 1 + \lambda_1 x_2 + \lambda_2(au_1 - bu_2).$$

The adjoint equations are

$$\begin{cases}\dot{\lambda}_1(t) = -H_{x_1} = 0,\\ \dot{\lambda}_2(t) = -H_{x_2} = -\lambda_1(t),\end{cases}$$

with the boundary conditions $\lambda_1(t_f) = \nu_1, \lambda_2(t_f) = \nu_2$. *The solutions of the adjoint system are easily obtained as*

$$\begin{cases}\lambda_1(t) = \nu_1,\\ \lambda_2(t) = \nu_2 - \nu_1(t - t_f).\end{cases}$$

Setting $\tau = \nu_2 - \nu_1(t - t_f)$ *and substituting* $\lambda_1(t)$ *and* $\lambda_2(t)$ *into* H*, we get*

$$H(t, x, u, \lambda) = 1 + \nu_1 x_2 + \tau(au_1 - bu_2) = 1 + \nu_1 x_2 + \tau a u_1 - \tau b u_2.$$

Since H *is linear in* u_1 *and* u_2*, it is simple to see that*

$$\min_{0 \le u_1, u_2 \le 1} H(t, x, u, \lambda) = \begin{cases} 1 + \nu_1 x_2 + 2\tau a \text{ if } \tau < 0 \\ 1 + \nu_1 x_2 - 2\tau b \text{ if } \tau \ge 0, \end{cases}$$

and the minimum is attained only if $u_1, u_2 \in \{0, 2\}$*. One can deduce that there is a time* $t_c < t_f$*, at which* $\tau = 0$*, such that an optimal control satisfies*

$$\begin{aligned} u_1(t) = 2, \quad u_2(t) = 0 \text{ if } t < t_c, \\ u_1(t) = 0, \quad u_2(t) = 2 \text{ if } t \ge t_c. \end{aligned}$$

In other words, the optimal controls are of bang-bang type. We can now compute the optimal solution effortlessly. We have for $t \in [0, t_c)$*,* $\dot{x}_2(t) = 2a$*. Hence*

$$x_2(t) = 2at, \quad x_1(t) = at^2.$$

For $t \in [t_c, t_f]$*,* $\dot{x}_2(t) = -2b$ *implies that*

$$\begin{aligned} x_2(t) &= 2at_c - 2b(t - t_c), \\ x_1(t) &= at_c^2 + 2at_c(t - t_c) - b(t - t_c)^2. \end{aligned}$$

Setting $x_2(t_f) = 0$ *and* $x_1(t_f) = 1$ *we obtain*

$$\frac{t_c}{t_f} = \frac{b}{a + b}, \quad t_f = \sqrt{\frac{a + b}{ab}}.$$

This solution is intuitively reasonable. If the ability to brake is high, i.e., b large, the car can accelerate in a long time, leading to reduction of the total time to reach the target. Finally, we would like to determine the adjoint variables. Using (2.9) *we get* $\nu_2 = 1/b$*. Setting* $\tau = 0$ *at* $t = t_c$ *we obtain* $\nu_1 = \nu_2/(t_c - t_f) = -(a + b)/(abt_f)$*.*

Example 2.3. *Consider a problem of moon-landing* (Evans [25])

$$\min_{u(\cdot), x(\cdot), t_f} -x_3(t_f) \tag{2.13}$$

subject to

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -g + \dfrac{u(t)}{x_3(t)} \\ \dot{x}_3(t) = -ku(t), \end{cases}$$

with the initial and terminal conditions

$$x_1(0) = h_0, \ x_2(0) = 0, \ x_3(0) = m_0; \quad x_1(t_f) = 0, \ x_2(t_f) = 0.$$

There are also constraints on the control, namely $0 \le u \le u_M$*. Here* $x_1(t)$ *represents the altitude,* $x_2(t)$ *the velocity and* $x_3(t)$ *the mass of the lander. Furthermore,* g, k, u_M, h_0, m_0 *are positive constants whose physical meaning is explained in Section 8.2.*

We are to show by means of the PMP that the optimal control is of bang-bang type. In fact, we have $\Phi(x) = -x_3,\ L(\cdot) = 0,\ \Psi(x) = (x_1, x_2)^T$. *Hence the Hamiltonian reads as*

$$H(t, x, u, \lambda) = \lambda_1 x_2 + \lambda_2(-g + u/x_3) - \lambda_3 k u.$$

The adjoint equation takes the form

$$\begin{cases} \dot{\lambda}_1(t) = -H_{x_1} = 0 \\ \dot{\lambda}_2(t) = -H_{x_2} = -\lambda_1(t) \\ \dot{\lambda}_3(t) = -H_{x_3} = \dfrac{u(t)}{x_3^2(t)}\lambda_2(t). \end{cases}$$

One can easily find that

$$\lambda_1(t) \equiv a,\ \lambda_2(t) = b - at,\ \dot{\lambda}_3(t) = u(t)(b - at)/x_3^2(t),$$

for some constants a, b. *Partially substituting those expressions into* H *yields*

$$H(t, x(t), u(t), \lambda(t)) = a x_2(t) - g\lambda_2(t) + u(t)\left(\frac{\lambda_2(t)}{x_3(t)} - k\lambda_3(t)\right).$$

Because H *is linear in* u, *the minimum of* H *with respect to* u *depends on the sign of* $A(t) = \lambda_2(t)/x_3(t) - k\lambda_3(t)$. *Taking the derivative of* $A(t)$ *with respect to* t *we obtain*

$$\begin{aligned} \dot{A}(t) &= \frac{\dot{\lambda}_2(t)}{x_3(t)} - \lambda_2(t)\frac{\dot{x}_3(t)}{x_3^2(t)} - k\dot{\lambda}_3(t) \\ &= -\frac{a}{x_3(t)} + \lambda_2(t)\frac{ku(t)}{x_3^2(t)} - k\frac{u(t)}{x_3^2(t)}\lambda_2(t) = -\frac{a}{x_3(t)}. \end{aligned}$$

Since $x_3(t)$ *represents the mass of the lander, it must be always positive. We deduce that* $\dot{A}(t)$ *does not change sign on* $[0, t_f]$. *As a consequence,* $A(t)$ *is monotone. Therefore, there exists* t_c, $0 \le t_c \le t_f$ *such that* $A(t) \ge 0$ *when* $0 \le t \le t_c$ *and* $A(t) \le 0$ *when* $t_c \le t \le t_f$ *(the reverse case leads to no solution). It follows from the linearity of* H *in* u *that the minimization of* H *corresponds to* $u(t) = 0$ *if* $0 \le t \le t_c$ *and* $u(t) = u_M$ *if* $t_c < t \le t_f$. *Once we know that an optimal control is of bang-bang type, we can derive a system of nonlinear equations to find* t_c *and* t_f. *Since the computation by hand is technical and is not essential, we do not go into details further. Numerical results are presented in Section* 8.2.

2.3 Hamilton-Jacobi-Bellman Equations and the sufficient conditions

In contrast to the PMP in which the effect of the variation of control on the objective function is considered, the Hamilton-Jacobi-Bellman Equations (HJBE) monitor the variation of the initial values and the initial time. The final time t_f is assumed to be fixed. In view of OCP (2.1)–(2.5), suppose that $V(t, x)$ is a function from $[t_0, t_f) \times \mathbb{R}^n$ into $\mathbb{R}$ that satisfies the HJBE

$$-V_t(t, x) = \min_{w \in \mathbb{U}} \Big\{ L(t, x, w) + V_x(t, x) f(t, x, w) \Big\}, \tag{2.14}$$

subject to the boundary condition

$$V(t_f, x) = \Phi(x) \text{ for all } x \in \big\{ x \in \mathbb{R}^{n_x} \mid \Psi(t_f, x) = 0 \big\}. \tag{2.15}$$

Theorem 2.2. (Zeidler [88, p. 85]) *Assume that there exists a once continuously differentiable function $V(t,x)$ satisfying* (2.14)–(2.15)*. If one can find a feasible control $u^*(\cdot)$ and the corresponding trajectory $x^*(\cdot)$ such that for all t at which $u^*(\cdot)$ is continuous, it holds that*

$$u^*(t) \in \operatorname{argmin}_{w\in\mathbb{U}} \Big\{ L(t,x^*(t),w) + V_x(t,x^*(t)) f(t,x^*(t),w) \Big\}, \tag{2.16}$$

then $u^*(\cdot)$ solves OCP (2.1)–(2.5) *with $V(t_0,x_0)$ as the optimal value.*

Remark 2.3. *The $u^*(\cdot)$ obtained from* (2.16) *has the form of a function of $(t,x^*(t))$. That means it defines a feedback law: The control action at a particular time is determined for the states of the system at that time. This is in sharp contrast to the PMP where the control is determined as a function of time and for the states only at the initial time. In feedback form, the control is able to react to changes in the system due to disturbances and various uncertainties which are often present in real-life processes.*

Remark 2.4. *If solutions of an OCP exist, one can define the value function $V(t,\bar{x})$ which is the optimal value of the OCP corresponding to the initial time t and the initial conditions $x(t) = \bar{x}$. This value function, however, may not satisfy the HJBE. The problem is that $V(t,\bar{x})$ may lack smoothness, especially for OCPs with constraints, as demonstrated in the following example.*

Example 2.4.

$$\min_{u(\cdot),x(\cdot)} x(1)$$

subject to

$$\begin{cases} \dot{x}(t) = x(t)u(t), \quad t \in [0,1], \\ x(t_0) = x_0, \end{cases}$$

$$-1 \le u(t) \le 1 \text{ for } 0 \le t \le 1.$$

Here $0 \le t_0 \le 1$ is an initial time.

Obviously, $x(t) = x_0 \exp\int_{t_0}^{t} u(\tau)d\tau$. Thus the value function has the form

$$V(t_0,x_0) = \begin{cases} x_0 \exp(-1+t_0) & \text{if } x_0 > 0, \\ x_0 \exp(1-t_0) & \text{if } x_0 < 0, \\ 0 & \text{if } x_0 = 0. \end{cases}$$

We have

$$\lim_{x_0\to 0^+} \frac{V(t_0,x_0) - V(t_0,0)}{x_0} = e^{-1+t_0};$$
$$\lim_{x_0\to 0^-} \frac{V(t_0,x_0) - V(t_0,0)}{x_0} = e^{1-t_0}.$$

Therefore $V(t_0,x_0)$ is not differentiable at $(t_0,0)$ for any $t_0 \neq 1$.

Hamilton-Jacobi-Bellman equation for the discrete case. Discrete-time systems account for an integral part of this work, not only because they are of interest in their own right but also because they represent sampling of continuous-time systems, as explained in Section 1.1. Consider a discrete-time system of the form

$$x_{k+1} = f_k(x_k, u_k), \quad k = 0, 1, 2, ... \tag{2.17}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^{n_u}$ and x_0 is given. The goal is to minimize with respect to $u = (u_0, u_1, ..., u_{N-1})$,

$$J_N(x_0, u) = F_N(x_N) + \sum_{k=0}^{N-1} L_k(x_k, u_k)$$

subject to (2.17). We introduce the value function

$$V(k, \bar{x}) = \inf_{u_k, u_{k+1}, ..., u_{N-1}} \left\{ F_N(x_N) + \sum_{i=k}^{N-1} L_i(x_i, u_i) \right\}, \quad k = 0, 1, ..., N,$$

with $x_k = \bar{x}$, $x_{i+1} = f_i(x_i, u_i)$, $i = k, k+1, ..., N-1$. The HJBE takes the form

$$V(k, \bar{x}) = \inf_{u_k} \Big\{ L_k(\bar{x}, u_k) + V(k+1, f_k(\bar{x}, u_k)) \Big\}, \quad k = N-1, N-2, ..., 0 \tag{2.18}$$

with $V_N(\bar{x}) = F_N(\bar{x})$. We can solve (2.18) backwards in k running from $N-1$ to 0 to obtain u_k in form of functions of x and then get x_k forward from (2.17). The optimal value is $V(0, x_0)$.

One of the remarkable applications of the HJBE is the *linear quadratic regulator* (LQR), in which the system is linear and the cost functions are quadratic, i.e.,

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k, \\ J_N(x_0, u) &= x_N^T Q_N x_N + \sum_{k=0}^{N-1} \left[x_k^T Q_k x_k + u_k^T R_k u_k \right], \end{aligned}$$

where $Q_k \succeq 0$, $R_k \succ 0$. The conditions $R_k \succ 0$ ensure that the cost function J_N is strictly convex with respect to u. Applying the HJBE and squares completion, we can prove by reverse induction that the value function is quadratic, concretely

$$V(k, \bar{x}) = \bar{x}^T S_k \bar{x}, \quad S_N = Q_N,$$

together with a recursive formula for S_k. In fact, it follows from (2.18) that

$$\begin{aligned} V(k, x_k) &= \min_{u_k} \Big\{ x_k^T Q_k (A_k x_k + B_k u_k) + u_k^T R_k u_k + V(k+1, A_k x_k + B_k u_k) \Big\} \\ &= \min_{u_k} \Big\{ u_k^T (R_k + B_k^T S_{k+1} B_k) u_k + 2 u_k^T B_k^T S_{k+1} A_k x_k + x_k^T (Q_k + A_k^T S_{k+1} A_k) x_k \Big\} \end{aligned}$$

The minimum is attained at

$$u_k^* = -(R_k + B_k^T S_{k+1} B_k)^{-1} B_k^T S_{k+1} A_k x_k,$$

and the minimal value is $V(k, x_k) = x_k^T S_k x_k$ where

$$S_k = Q_k + A_k^T S_{k+1} A_k - A_k^T S_{k+1} B_k (R_k + B_k^T S_{k+1} B_k)^{-1} B_k^T S_{k+1} A_k.$$

Those follow from an elementary lemma on the minimization of quadratic functions,

$$\min_w \left\{ w^T R w + 2 w^T H \bar{x} + \bar{x}^T Q \bar{x} \right\} = \bar{x}^T (Q - H^T R^{-1} H) \bar{x}, \tag{2.19}$$

where $R \succ 0$. The minimum is attained at $w^* = -R^{-1} H \bar{x}$.

2.4 Sensitivity analysis of OCPs

2.4.1 OCPs under uncertainty

In this section, we derive sensitivity formulas for OCPs *without constraints*. Concretely, we consider OCP (2.1)–(2.3) and deal with parameters p explicitly, i.e.,

$$\min_{u,x} J(x,u) = \Phi(x(t_f)) + \int_{t_0}^{t_f} L(t,x(t),u(t))dt, \tag{2.20}$$

subject to

$$\begin{cases} \dot{x}(t) = f(t,x(t),u(t),p), \quad t \in [t_0,t_f], \\ x(t_0) = x_0, \end{cases} \tag{2.21}$$

We assume that x_0 is uncertain and generally, x_0 depends on p, i.e., $x_0 = x_0(p)$. Note that for a concrete OCP, we fix the parameters at their nominal values $p = p_0$.

2.4.2 Sensitivity analysis of OCPs in the framework of PMP

If the parameters are unknown, it is desirable to know how much the optimal values or the solutions change when the parameters vary. In this section we are interested in local sensitivity of the optimal value, which is explored by its derivatives with respect to the parameters. To this end, let $J^*(p)$ denote the optimal value of (2.20) corresponding to p. Our goal is to calculate the derivative $\frac{\mathrm{d}J^*}{\mathrm{d}p}(p)$.

For simplicity of notation, we often omit arguments of functions without causing any misunderstanding. Let p be fixed and $x^*(\cdot), u^*(\cdot)$ solve OCP (2.20)-(2.21). The adjoint equations according to (2.7) read as

$$\begin{cases} \dot{\lambda}^T(t) = -\lambda^T(t) f_x(t,x^*,u^*,p) - L_x(t,x^*,u^*), \ t \in (0,t_f) \\ \lambda(t_f) = \Phi_x^T(x^*(t_f)). \end{cases}$$

Since there are no constraints, it follows from (2.10) that

$$\lambda^T f_u + L_u = 0.$$

We suppose that $x^*(\cdot), u^*(\cdot), \lambda(\cdot)$ as well as the optimal value $J^*(p)$ are differentiable with respect to p. The derivative $\frac{\mathrm{d}J^*}{\mathrm{d}p}(p)$ is closely related to the Lagrange multipliers as shown in the following lemma.

Lemma 2.1. *The following equality is valid*

$$\frac{\mathrm{d}J^*}{\mathrm{d}p}(p) = \lambda^T(t_0)\frac{\partial x^*(t_0)}{\partial p} + \int_{t_0}^{t_f} \lambda^T(t)\frac{\partial f}{\partial p}(t,x^*(t),u^*(t),p)dt.$$

Proof. By the definition of $J^*(p)$, we have

$$\begin{aligned} \frac{\mathrm{d}J^*}{\mathrm{d}p}(p) &= \frac{\partial \Phi}{\partial x}(x^*(t_f))\frac{\partial x^*(t_f)}{\partial p} + \int_{t_0}^{t_f} \left(\frac{\partial L}{\partial x}\frac{\partial x^*}{\partial p} + \frac{\partial L}{\partial u}\frac{\partial u^*}{\partial p}\right) dt \\ &= \lambda^T(t_f)\frac{\partial x^*(t_f)}{\partial p} + \int_{t_0}^{t_f} \left(\frac{\partial L}{\partial x}\frac{\partial x^*}{\partial p} + \frac{\partial L}{\partial u}\frac{\partial u^*}{\partial p}\right) dt. \end{aligned} \tag{2.22}$$

On the other hand

$$\begin{aligned}\lambda^T(t_f)\frac{\partial x^*(t_f)}{\partial p} =&\lambda^T(t_0)\frac{\partial x^*(t_0)}{\partial p} + \int_{t_0}^{t_f} \frac{d}{dt}\left(\lambda^T(t)\frac{\partial x^*}{\partial p}(t)\right) dt \\ =&\lambda^T(t_0)\frac{\partial x^*(t_0)}{\partial p} + \int_{t_0}^{t_f} \dot{\lambda}^T(t)\frac{\partial x^*}{\partial p}(t)dt + \int_{t_0}^{t_f} \lambda^T(t)\frac{\partial \dot{x}^*}{\partial p}(t)dt \\ =&\lambda^T(t_0)\frac{\partial x^*(t_0)}{\partial p} + \int_{t_0}^{t_f} \dot{\lambda}^T(t)\frac{\partial x^*}{\partial p}(t)dt \\ &+ \int_{t_0}^{t_f} \lambda^T(t)\left(\frac{\partial f}{\partial x}\frac{\partial x^*}{\partial p} + \frac{\partial f}{\partial p} + \frac{\partial f}{\partial u}\frac{\partial u^*}{\partial p}\right) dt.\end{aligned}$$

Substituting the last equation into (2.22), we obtain

$$\begin{aligned}\frac{\mathrm{d}J^*}{\mathrm{d}p}(p) =&\lambda^T(t_0)\frac{\partial x^*(t_0)}{\partial p} + \int_{t_0}^{t_f} \left(\dot{\lambda}^T(t) + \frac{\partial L}{\partial x} + \lambda^T(t)\frac{\partial f}{\partial x}\right)\frac{\partial x^*}{\partial p}dt \\ &+ \int_{t_0}^{t_f} \left(\lambda^T(t)\frac{\partial f}{\partial u} + \frac{\partial L}{\partial u}\right)\frac{\partial u^*}{\partial p}dt + \int_{t_0}^{t_f} \lambda^T(t)\frac{\partial f}{\partial p}dt \\ =&\lambda^T(t_0)\frac{\partial x^*(t_0)}{\partial p} + \int_{t_0}^{t_f} \lambda^T(t)\frac{\partial f}{\partial p}(x^*(t), u^*(t), p)dt\end{aligned}$$

since

$$\dot{\lambda}^T(t) + \frac{\partial L}{\partial x} + \lambda^T(t)\frac{\partial f}{\partial x} = 0, \quad \lambda^T(t)\frac{\partial f}{\partial u} + \frac{\partial L}{\partial u} = 0.$$

The proof is complete. □

Remark 2.5. *In a special case when $x(t_0) = p$ and f does not depend explicitly on p, i.e., $f_p = 0$, we have*

$$\frac{\mathrm{d}J^*}{\mathrm{d}p}(p) = \lambda^T(t_0)\frac{\partial x^*(t_0)}{\partial p} = \lambda^T(t_0).$$

In solving OCPs numerically, we often deal with nonlinear programs (NLPs) of the form

$$\begin{cases}\min & \psi(\mathbf{x}) \\ \text{s.t} & g(\mathbf{x}, p) = 0.\end{cases} \tag{2.23}$$

where $\mathbf{x} \in \mathbb{R}^N$, $p \in \mathbb{R}^{n_p}$, and $g : \mathbb{R}^N \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_g}$. Assume that optimal value $\psi^*(p)$ and the solution $\mathbf{x}^*(p)$ are differentiable with respect to p.

Lemma 2.2. *Let $\lambda(p)$ be the corresponding Lagrange multiplier of (*2.23), then

$$\frac{\mathrm{d}\psi^*}{\mathrm{d}p}(p) = -\lambda^T(p)\frac{\partial g}{\partial p}(\mathbf{x}^*(p), p).$$

Proof. The Lagrange function reads as

$$\mathbf{L}(\mathbf{x}, p, \lambda) = \psi(\mathbf{x}) - \lambda^T(p)g(\mathbf{x}, p).$$

According to the *Karush-Kuhn-Tucker* (KKT) optimality conditions

$$\frac{\partial \psi}{\partial \mathbf{x}}(\mathbf{x}^*(p)) = \lambda^T(p)\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}^*(p), p).$$

Multiplying both sides of this equality with $\frac{\partial \mathbf{x}^*}{\partial p}(p)$, we obtain

$$\frac{\partial \psi}{\partial \mathbf{x}}(\mathbf{x}^*(p))\frac{\partial \mathbf{x}^*}{\partial p}(p) = \lambda^T(p)\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}^*(p),p)\frac{\partial \mathbf{x}^*}{\partial p}(p).$$

On the other hand, differentiating the side condition $g(\mathbf{x},p)=0$ with respect to p yields

$$\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}^*(p),p)\frac{\partial \mathbf{x}^*}{\partial p}(p) + \frac{\partial g}{\partial p}(\mathbf{x}^*(p),(p)) = 0.$$

Combining the last three equalities we obtain

$$\frac{\mathrm{d}\psi^*}{\mathrm{d}p}(p) = \frac{\partial \psi}{\partial \mathbf{x}}(\mathbf{x}^*(p))\frac{\partial \mathbf{x}^*}{\partial p}(p) = -\lambda^T(p)\frac{\partial g}{\partial p}(\mathbf{x}^*(p),p).$$

This completes the proof. □

Remark 2.6. *Lemma* 2.2 *can be generalized to NLPs with inequality constraints. The idea is to make use of complementarity conditions and treat the nonlinear program at hand locally as ones with only equality constraints.*

Remark 2.7. *From a computational point of view, it might be convenient to calculate* $\frac{\mathrm{d}\psi^*}{\mathrm{d}p}(p_0)$ *in the nonlinear program* (2.23) *as follows*

- *Introduce new variables* p *and additional constraints* $p - p_0 = 0$
- *Solve the modified nonlinear program*
- *Then* $\frac{\mathrm{d}\psi^*}{\mathrm{d}p}(p_0)$ *is equal to the Lagrange multiplier with respect to the constraint* $p - p_0 = 0$.

We will pursue this strategy in the numerical implementation of Dual Control methods presented in Chapter 6.

2.4.3 Sensitivity analysis in the framework of HJBE

The HJBE is concerned with the derivatives of the value function $V(t,x)$ which are helpful for sensitivity analysis of OCPs with respect to the initial time and the initial values. Using the *envelope theorem*, Appendix A, one can prove that if solutions of OCP (2.1)–(2.4) exist, and if the value function satisfies the HJBE, then the adjoint variables $\lambda(t)$ in the PMP possess an interesting property:

$$\lambda^T(t_0) = V_x(t_0,x_0),$$

(cf. Bertsekas [10]). To prove that, let us first recall the HJBE

$$-V_t(t,x) = \min_{w\in\mathbb{U}}\Big\{L(t,x,w) + V_x(t,x)f(t,x,w)\Big\}, \tag{2.24}$$

Denote by $u^0 = u^0(t,x)$ a minimizer of the minimization problem in (2.24). We do not require the uniqueness of solutions. It follows from Theorem A.1, Appendix A that

$$-V_{tx}(t,x) = L_x(t,x,u^0) + V_x(t,x)f_x(t,x,u^0) + V_{xx}(t,x)f(t,x,u^0),$$

and

$$-V_{tt}(t,x) = L_t(t,x,u^0) + V_x(t,x)f_t(t,x,u^0) + V_{tx}(t,x)f(t,x,u^0).$$

We consider these equalities along solutions $x^*(t), u^*(t)$ of the OCP. It holds that

$$\begin{aligned}
\frac{d}{dt}V_x(t,x^*(t)) =& V_{xx}(t,x^*(t))\dot{x}^*(t) + V_{tx}(t,x^*(t)) \\
=& V_{xx}(t,x^*(t))f(t,x^*(t),u^*(t)) + V_{tx}(t,x^*(t)) \\
=& V_{xx}(t,x^*(t))f(t,x^*(t),u^*(t)) - \Big[L_x(t,x^*(t),u^*(t)) \\
& + V_x(t,x^*(t))f_x(t,x^*(t),u^*(t)) + V_{xx}(t,x^*(t))f(t,x^*(t),u^*(t))\Big] \\
=& - L_x(t,x^*(t),u^*(t)) - V_x(t,x^*(t))f_x(t,x^*(t),u^*(t)).
\end{aligned}$$

and

$$\begin{aligned}
\frac{d}{dt}V_t(t,x^*(t)) =& V_{tx}(t,x^*(t))\dot{x}^*(t) + V_{tt}(t,x^*(t)) \\
=& V_{tx}(t,x^*(t))f(t,x^*(t),u^*(t)) + V_{tt}(t,x^*(t)) \\
=& V_{tx}(t,x^*(t))f(t,x^*(t),u^*(t)) - \Big[L_t(t,x^*(t),u^*(t)) \\
& + V_x(t,x^*(t))f_t(t,x^*(t),u^*(t)) + V_{tx}(t,x^*(t))f(t,x^*(t),u^*(t))\Big] \\
=& - L_t(t,x^*(t),u^*(t)) - V_x(t,x^*(t))f_t(t,x^*(t),u^*(t)).
\end{aligned}$$

for all $t \in (t_0,t_f)$. By setting $\lambda(t) = (V_x(t,x^*(t)))^T$ and $p_0(t) = V_t(t,x^*(t))$, we can simplify these expressions above as

$$\begin{aligned}
\dot{\lambda}^T(t) &= -L_x(t,x^*(t),u^*(t)) - \lambda^T(t)f_x(t,x^*(t),u^*(t)), \\
\dot{p_0}(t) &= -L_t(t,x^*(t),u^*(t)) - \lambda^T(t)f_t(t,x^*(t),u^*(t)).
\end{aligned}$$

We thus obtain the adjoint equations as in the PMP. The boundary conditions for $\lambda(t)$ may be derived using Lemma A.1, Appendix A. In fact, since $V(t_f,x) = \Phi(x)$ for all $x \in K = \{x \mid \Psi(t_f,x) = 0\}$, there exists $\nu \in \mathbb{R}^{n_\Psi}$ such that

$$V_x(t_f,x) = \Phi_x(x) + \nu^T\Psi_x(t_f,x), \quad x \in K.$$

Therefore

$$\lambda^T(t_f) = \Phi_x(x^*(t_f)) + \nu^T\Psi_x(t_f,x^*(t_f)).$$

At $t = 0$, we have

$$V_x(t_0,x_0) = \lambda^T(t_0),$$

which is the derivative of the optimal value with respect to the initial values x_0. This was previously shown in the context of the PMP by Remark 2.5. Thus we have one more way to derive the sensitivity of OCPs.

Chapter 3

Asymptotic Properties of Nonlinear Least Squares Estimates

The least squares estimator for multivariate nonlinear regression with independent, identically distributed data is well-known to be consistent under the condition of uniform observability and to be asymptotically normally distributed. For dynamic systems, the condition of uniform observability may not hold and we show that, for large sample sizes, the estimation problem may become ill-posed, leading to inconsistency of the least squares estimator. However, the sequential least squares method when combined with regularization can retain convergence and a well-behaved asymptotic distribution, and these results are illustrated by several examples. We generalize to the estimation of parameters in dynamic systems the Cramér-Rao and Bhattacharya inequalities. Further, we provide an algorithm for applying the sequential least squares strategy, and we prove a local convergence result for the algorithm.

3.1 Introduction

In the statistics literature, the method of *maximum likelihood estimation* is widely applied because of its attractive theoretical properties. When measurements are obtained from independent, identically distributed (i.i.d.) random variables, the consistency and asymptotic normality of maximum likelihood estimators hold under general conditions. Suppose that the common probability density function of these random variables is $f(x;p)$, where $x \in \mathbb{R}^n$ is an observed value of the underlying multivariate random vector and $p \in \mathbb{R}^{n_p}$ is an unknown parameter. Denote also by p^* the true value of the parameter p. An estimator $\hat{p}^N_{\mathrm{MLE}}$ which is based on a random sample of size N and which maximizes the likelihood function is called a *maximum likelihood estimator* (MLE). Under certain regularity conditions on $f(x;p)$, it is well-known that $\hat{p}^N_{\mathrm{MLE}}$ satisfies the properties of:

- **Strong consistency:** $\hat{p}^N_{\mathrm{MLE}} \xrightarrow{\text{a.s}} p^*$ almost surely (a.s.), and therefore also satisfies **Consistency:** $\hat{p}^N_{\mathrm{MLE}} \xrightarrow{\mathbb{P}} p^*$ in probability as $N \to \infty$.

- **Asymptotic normality:** $\sqrt{N}(\hat{p}^N_{\mathrm{MLE}} - p^*) \xrightarrow{\mathrm{d}} \mathcal{N}(0, C)$, i.e., convergence in distribution to $\mathcal{N}(0, C)$, a multivariate normal distribution with mean vector 0 and an appropriately chosen covariance matrix C.

Moreover, $\mathrm{Cov}(\hat{p}^N_{\mathrm{MLE}} - p^*)$ converges to the inverse of the Fisher information matrix, proving that the MLE is asymptotically optimal (Bahadur [6]).

In the case of Gaussian noise, the *least squares* (LS) estimator and the MLE coincide (see Section 3.2), and this raises the issue of the asymptotic properties of LS estimators in general. Indeed, the consistency and asymptotic normality of LS estimators hold under general assumptions on the distribution of the data, and no Gaussian assumptions are needed; see Ivanov [39, 40], Jennrich [43], and Prakasa Rao [70].

The results stated for the above *static* case are relatively straightforward because measurements are uniformly informative. That is, each measurement provides the same amount of information as measured by the information matrix defined in Section 3.2; this leads to a build-up of information, and results in convergence of the estimator.

In the case of nonstationary processes, the asymptotic behavior of the parameter estimators can be different. In this setting, low signal-to-noise ratio or information die-off may cause theoretical and computational difficulties. These problems may also result in a loss of consistency, as demonstrated by the following example.

Example 3.1. Consider the regression model,

$$y_t = \frac{p}{t} + \varepsilon_t, \qquad t = 1, 2, \ldots \tag{3.1}$$

where the random variables y_t and the constant p are scalars, and the ε_t are mutually independent, random variables with mean 0 and variance 1. For this model $\hat{p}^N_{\mathrm{LS}}$, the least squares estimator of p, has an explicit form. Defining $S_N^2 = \sum_{t=1}^N t^{-2}$, we have

$$\hat{p}^N_{\mathrm{LS}} = S_N^{-2} \sum_{t=1}^{N} \frac{y_t}{t}.$$

Further, $\mathrm{Var}(\hat{p}^N_{\mathrm{LS}}) = S_N^{-2}$; and since $\lim_{N\to\infty} S_N^{-2} = 6/\pi^2 > 0$ then $\hat{p}^N_{\mathrm{LS}}$ is not consistent. Depending on the distributional characteristics of the noise variables ε_t, we can also ascertain the asymptotic properties of $\hat{p}^N_{\mathrm{LS}}$. For example, if the ε_t are all Gaussian white noise then we can show that $(\hat{p}^N_{\mathrm{LS}} - p^*) \xrightarrow{\mathrm{d}} \mathcal{N}(0, 6/\pi^2)$.

On the other hand, if the ε_t are non-Gaussian then the limiting distribution, if it exists, generally is more complicated; for instance, if the ε_t are uniformly distributed then $\hat{p}^N_{\mathrm{LS}} - p^*$ is not asymptotically Gaussian. These examples indicate also that the asymptotic behavior of least squares estimators in the case of inconsistency depends on the distribution of the noise variables. We shall provide in Example 3.2 further details for this case and relate more general formulations of the regression model (3.1) to the subject of random walks with variable-length step-sizes.

Because the system in Example 3.1 is linear in p, the estimator for each sample size is unique and has an explicit form. In contrast, nonlinear models are well-known to exhibit numerous theoretical and numerical difficulties. For instance, the solution may have no closed form; or it may not be unique, in which case the existence of local optima may complicate the relationship between computed solutions and the distribution of the noise errors; cf. Seber and Wild [79, Chapter 3].

We shall, by analyzing the least squares objective function, explain the ill-posedness of the least squares estimation problem when the sample size is large. It is shown that, for some models, the least squares estimator is not consistent, and the least squares objective function for large sample sizes is asymptotically independent of the parameters.

Conventional numerical methods can lead to non-convergence of estimates even when theoretical results prove consistency. To resolve these difficulties, we propose a sequential least squares method which processes data stepwise in moving time intervals. This sequential methods helps to ensure convergence of the computed estimates in the case of consistency. Moreover, in cases which lack consistency, the sequential least squares methods is powerful enough to provide computed estimates that are well-distributed around the true value of the parameters.

The chapter is organized as follows. In Section 3.2, we generalize the well-known Cramér-Rao and Bhattacharya inequalities to dynamic systems. Section 3.3 summarizes some results in the literature on the consistency of least squares estimators and analyzes for some crucial examples the asymptotic properties of least squares estimators by means of the behavior of the corresponding least squares objective functions. Section 3.4 introduces the sequential least squares method, describes the results of simulations for some illustrative numerical examples, and provides a brief comment on Gauss' work on the discovery of a celestial object, Ceres, and the connections between his work and least squares estimation and the sequential least squares method. Finally, in Section 3.5, we provide an algorithm for applying the sequential least squares strategy, and establish a local convergence result for the algorithm.

3.2 Maximum likelihood and least squares estimation for dynamic systems

The system under consideration is described by the statistical regression model

$$y_t = h(x_t; p) + \varepsilon_t, \quad t = 1, 2, \ldots \tag{3.2}$$

where the $x_t \in \mathbb{R}^n$ are called the regression variables; $p \in \mathbb{R}^{n_p}$ are the unknown parameters; the data $y_t \in \mathbb{R}^m$ are noisy measurements; the unobservable random noise variables $\varepsilon_t \in \mathbb{R}^m$; and $h(x_t; p) \in \mathbb{R}^m$ is the regression function.

We assume that $p \in \Theta$, a compact subset of $\mathbb{R}^{n_p}$; that the regression function $h(x_t; p)$ is continuous in x_t and continuously differentiable with respect to p; and that the ε_t are i.i.d. and have finite second moments. For brevity, we use throughout the paper the notation

$$h_t(p) \equiv h(x_t; p).$$

Model (3.2) can be regarded as an output model for a time-varying system with states $x(t) = x_t$ depending on unknown parameters p. In order to estimate p, we collect measurements at the sampling time points t on the regression function $h_t(p)$, recording the measurements y_t which have been masked by the noise ε_t.

Suppose that we have measurements $y_1, \ldots, y_N$ which are realizations of independent random variables $Y_1, \ldots, Y_N$. Our task is to use these measurements to estimate p by means of the model (3.2).

Let $y^N = (y_1, \ldots, y_N) \in \mathbb{R}^{m \times N}$, the space of $m \times N$ matrices, and let $Y^N = (Y_1, \ldots, Y_N)$. If each Y_t has the density function $f_t(y_t; p)$ then the density function of Y^N is given by

$$f(y^N; p) = \prod_{t=1}^{N} f_t(y_t; p). \tag{3.3}$$

In particular, if each $\varepsilon_t \sim \mathcal{N}(0, R_t)$ where $R_t \succ 0$, then

$$f(y^N; p) = \frac{1}{(2\pi)^{mN/2} \prod_{t=1}^{N} (\det R_t)^{1/2}} \exp\Big[-\tfrac{1}{2} \sum_{t=1}^{N} \big(y_t - h_t(p)\big)^T R_t^{-1} \big(y_t - h_t(p)\big) \Big]. \quad (3.4)$$

In what follows we make the assumption that $f(\cdot)$ is regular enough such that the interchangeability of integration and differentiation is applicable. Similar to the theory of estimation for static processes, we obtain the following results.

Lemma 3.1. *Under the assumption that the observations Y_t, $t = 1, 2, \ldots$ are mutually independent,*

$$\mathbb{E}\Big(\frac{\partial}{\partial p} \log f(Y^N; p) \Big) = 0.$$

Proof. Using the density function (3.3), we obtain

$$\begin{aligned} \mathbb{E}\Big(\frac{\partial}{\partial p} \log f(Y^N; p) \Big) &= \int_{\mathbb{R}^{m \times N}} f(y^N; p) \Big(\frac{\partial}{\partial p} \log f(y^N; p) \Big) \, \mathrm{d}y^N \\ &= \int_{\mathbb{R}^{m \times N}} \frac{\partial}{\partial p} f(y^N; p) \, \mathrm{d}y^N \\ &= \frac{\partial}{\partial p} \int_{\mathbb{R}^{m \times N}} f(y^N; p) \, \mathrm{d}y^N = \frac{\partial}{\partial p}(1) = 0, \end{aligned}$$

which completes the proof. □

Define the $p \times 1$ column vector

$$B(y^N; p) = \frac{\partial}{\partial p} \log f(y^N; p),$$

and the $p \times p$ matrix

$$\begin{aligned} M_N(p) &= \mathbb{E}\big(B(Y^N; p)\, B^T(Y^N; p) \big) \\ &= \int_{\mathbb{R}^{m \times N}} B(y^N; p) B^T(y^N; p) f(y^N; p) \, \mathrm{d}y^N. \end{aligned}$$

In the static case, the matrix $M_N(p)$ is the well-known *(Fisher) information matrix.*

Lemma 3.2. *The information matrix is given by*

$$M_N(p) = -\mathbb{E}\Big[\Big(\frac{\partial}{\partial p} \Big) \Big(\frac{\partial}{\partial p} \Big)^T \log f(Y^N; p) \Big].$$

The proof of this identity is obtained by differentiating both sides of the equality of Lemma 3.1. As usual, an estimator $\hat{p}^N \equiv \hat{p}^N(Y^N)$ is said to be *unbiased* if $\mathbb{E}(\hat{p}^N) = p$ for all p, and then its covariance matrix is $\mathrm{Cov}(\hat{p}^N) := \mathbb{E}\big(\hat{p}^N(Y^N) - p\big)\big(\hat{p}^N(Y^N) - p\big)^T$.

Lemma 3.3. (Generalized Cramér-Rao inequality) *Suppose that the information matrix $M_N(p)$ is nonsingular. For any unbiased estimator $\hat{p}^N(Y^N)$ of p, there holds the inequality*

$$\mathrm{Cov}(\hat{p}^N) \succeq M_N^{-1}(p). \quad (3.5)$$

Proof. By the unbiasedness of $\hat{p}^N$,

$$\int_{\mathbb{R}^{m\times N}} \hat{p}^N(y^N) f(y^N;p)\,\mathrm{d}y^N \equiv \mathbb{E}(\hat{p}^N) = p.$$

Differentiating this equality with respect to p we obtain

$$\int_{\mathbb{R}^{m\times N}} \hat{p}^N(y^N) \left(\frac{\partial}{\partial p} f(y^N;p)\right) \mathrm{d}y^N = I_m;$$

equivalently,

$$\int_{\mathbb{R}^{m\times N}} \hat{p}^N(y^N) \left(\frac{\partial}{\partial p} \log f(y^N;p)\right) f(y^N;p)\,\mathrm{d}y^N = I_m.$$

Define $d\mu(y^N) = f(y^N;p)\,\mathrm{d}y^N$, which is a probability measure on $\mathbb{R}^{m\times N}$. Since

$$\int_{\mathbb{R}^{m\times N}} B(y^N;p)\,\mathrm{d}\mu(y^N) = 0$$

then it follows that

$$\int_{\mathbb{R}^{m\times N}} (\hat{p}^N - p)B^T(y^N;p)\,\mathrm{d}\mu(y^N) = I_m.$$

For brevity, set $A(y^N) = \hat{p}^N - p$ and $B(y^N) = B(y^N;p)$. To complete the proof, we need to show that if $\int_{\mathbb{R}^{m\times N}} A(y)B^T(y)\,\mathrm{d}\mu(y) = I_m$ and $\int_{\mathbb{R}^{m\times N}} B(y)B^T(y)\,\mathrm{d}\mu(y)$ is nonsingular then

$$\int_{\mathbb{R}^{m\times N}} A(y)A^T(y)\,\mathrm{d}\mu(y) \succeq \left(\int_{\mathbb{R}^{m\times N}} B(y)B^T(y)\,\mathrm{d}\mu(y)\right)^{-1}.$$

This result follows from the more general lemma given below. □

Lemma 3.4. *Suppose that $A(y)$ and $B(y)$ are real-valued matrix functions from a measurable space (Ω,μ) to $\mathbb{R}^{m\times N}$ such that $\int_\Omega B(y)B^T(y)\,d\mu(y)$ is nonsingular. Then,*

$$\int_\Omega A(y)A^T(y)\,d\mu(y) \succeq \left(\int_\Omega A(y)B^T(y)\,d\mu(y)\right)\left(\int_\Omega B(y)B^T(y)\,d\mu(y)\right)^{-1}\left(\int_\Omega B(y)A^T(y)\,d\mu(y)\right).$$

Proof. For every $y \in \Omega$,

$$\begin{bmatrix} A(y)A^T(y) & A(y)B^T(y) \\ B(y)A^T(y) & B(y)B^T(y) \end{bmatrix} \equiv \begin{bmatrix} A(y) \\ B(y) \end{bmatrix} \begin{bmatrix} A^T(y) & B^T(y) \end{bmatrix} \succeq 0.$$

Integrating this inequality over Ω, we deduce that

$$\begin{bmatrix} \int_\Omega A(y)A^T(y)\,\mathrm{d}\mu(y) & \int_\Omega A(y)B^T(y)\,\mathrm{d}\mu(y) \\ \int_\Omega B(y)A^T(y)\,\mathrm{d}\mu(y) & \int_\Omega B(y)B^T(y)\,\mathrm{d}\mu(y) \end{bmatrix} \equiv \int_\Omega \begin{bmatrix} A(y)A^T(y) & A(y)B^T(y) \\ B(y)A^T(y) & B(y)B^T(y) \end{bmatrix} \mathrm{d}\mu(y) \succeq 0.$$

Applying the Schur complement lemma yields the desired conclusion. □

In the classical static case, Lemma 3.3 is generalized by the Bhattacharya inequality (see Bahadur [6, Chapter 4]). We now extend this inequality to the nonstatic setting.

Lemma 3.5. (Generalized Bhattacharya inequality) *Suppose that the function $g(p)$ is twice continuously differentiable in p, and let $M_N(p)$ be nonsingular. Denote by $J_g(p)$ the Jacobian of g with respect to p. For any unbiased estimator $\hat{\tau}(Y^N)$ of $g(p)$, there holds the inequality,*

$$\mathrm{Cov}\left(\hat{\tau}(Y^N)\right) \succeq J_g(p) M_N^{-1}(p) J_g(p)^T. \tag{3.6}$$

The proof of this result is obtained similarly to the proof of Lemma 3.3. We now consider a special but common case in which the measurement errors are assumed to be white noise, i.e., $\varepsilon_t \sim \mathcal{N}(0, R_t)$, where $R_t \succ 0$ for all t. The density function of Y^N is given in (3.4), and it follows that the log-likelihood function is

$$l_N(y^N; p) = -\frac{1}{2}\sum_{t=1}^{N} \log((2\pi)^m \det(R_t)) - \frac{1}{2}\sum_{t=1}^{N} (y_t - h_t(p))^T R_t^{-1} (y_t - h_t(p)),$$

and also that

$$\begin{aligned} B(y^N; p) &= \frac{\partial^T}{\partial p} l_N(y^N; p) \\ &= \sum_{t=1}^{N} \frac{\partial h_t(p)}{\partial p}^T R_t^{-1} (y_t - h_t(p). \end{aligned}$$

Note that $\mathbb{E}(Y_s - h_s(p))(Y_t - h_t(p))^T = \delta_{st} R_t$, where δ_{st} denotes Kronecker's delta; i.e., $\delta_{st} = 1$ if $s = t$ and $\delta_{st} = 0$, otherwise. The information matrix therefore takes the form,

$$\begin{aligned} M_N(p) &= \mathbb{E}\left(B(Y^N; p) B^T(Y^N; p)\right) \\ &= \sum_{t=1}^{N} \Big(\frac{\partial h_t(p)}{\partial p}\Big)^T R_t^{-1} \frac{\partial h_t(p)}{\partial p}. \end{aligned}$$

We define the *least squares objective function*,

$$L_N(y^N; p) = \frac{1}{2N}\sum_{t=1}^{N} (y_t - h_t(p))^T R_t^{-1} (y_t - h_t(p)).$$

Consequently, we see that the MLEs coincide with the least squares estimators.

If the regression function $h_t(p)$ is linear in p and the noise is white then the MLE is unbiased and satisfies the equality in Lemma 3.3. In general, however, it is difficult to determine whether an MLE or least squares estimator is unbiased or achieves the Cramér-Rao lower bound. Results in the cases of static parameter estimation state that under certain conditions, the MLE is asymptotically unbiased and asymptotically optimal in the sense that its covariance matrix converges to the Cramér-Rao lower bound as $N \to \infty$.

On the other hand, the inverse of the information matrix can be interpreted as a linear approximation of the covariance matrix of the solution to the problem of minimizing the least squares function; cf., Bock [11], Körkel [46]. This is the standpoint for the use of the MLE and the least squares estimators in practice. Moreover, this approximation has served as the basis for the construction of optimal experimental designs which have been successfully applied to chemical processes (Körkel [46]).

The following lemma provides a useful necessary condition, in terms of the information matrix, for the consistency of an estimator.

Lemma 3.6. *Let $\hat{p}^N$ be a sequence of unbiased estimators of $p \in \Theta$. If*

$$\limsup_{N\to\infty} \min_{\tilde{p}\in\Theta} \|M_N^{-1}(\tilde{p})\| > 0, \tag{3.7}$$

where $\|\cdot\|$ is any matrix norm, then $\hat{p}^N$ is not consistent.

Proof. Suppose, to the contrary, that $\hat{p}^N$ is consistent. Then $\mathrm{Cov}(\hat{p}^N)$, the covariance matrix of the estimator, satisfies

$$\lim_{N\to\infty} \|\operatorname{Cov}(\hat{p}^N)\| = 0.$$

By Lemma 3.3,

$$\operatorname{Cov}(\hat{p}^N) \succeq \min_{\tilde{p}\in\Theta} M_N^{-1}(\tilde{p}).$$

Then it follows that

$$\limsup_{N\to\infty} \min_{\tilde{p}\in\Theta} M_N^{-1}(\tilde{p}) = 0,$$

which contradicts (3.7). The proof is complete. □

3.3 Asymptotic behavior of the least squares estimators

In the case of finitely many measurements, it may be difficult to determine the statistical properties of an estimator. As the number of measurements increases, however, it is natural to hope that the sequence of estimators will converge to the true value of the parameter and also converge in distribution. This question can be answered satisfactorily in the case of static models with i.i.d. measurements. In this case the MLEs, and therefore the least squares estimators in the case of white noise, are asymptotically unbiased and asymptotically optimal with respect to the Cramér-Rao inequality.

Let us now consider model (3.2). Since x_t varies with t, measurements y_t are not i.i.d. While, loosely speaking, the information in the static case is constant and increases when new measurements are obtained, the information in the nonstatic case might die off or decay gradually with time. These issues were investigated intensively both in the statistical and control theory literature, and several general results are available.

Under the condition of *uniform observability*, which we give in (3.8) below, the least squares estimators are convergent in probability and even almost surely (a.e.) with specified rates of convergence (linear or exponential). Moreover, the distribution of these estimators can be asymptotically normal even if the measurement noise is not Gaussian. Assuming, without loss of generality, that measurements are scalar, we consider the function

$$\phi_N(p_1, p_2) = \frac{1}{2N} \sum_{t=1}^{N} \big(h_t(p_1) - h_t(p_2)\big)^2.$$

Denote by $\hat{p}^N$ the least squares estimator of the parameter p^* based on N measurements.

Theorem 3.1. (Ivanov [39, 40]) *Suppose that there exist $\alpha_1, \alpha_2 > 0$ such that*

$$\alpha_1 |p_1 - p_2|^2 \le \phi_N(p_1, p_2) \le \alpha_2 |p_1 - p_2|^2 \tag{3.8}$$

for all $p_1, p_2 \in \Theta$ and $N > 0$. Then there exists a constant $c > 0$ such that for all sufficiently large N and all $\varepsilon > 0$,

$$\mathbb{P}(\sqrt{N}(\hat{p}^N - p^*) > \varepsilon) < \frac{c}{\varepsilon^2}.$$

Moreover, we have an exponential rate of convergence in this context if the noise distributions are Gaussian. We remark that the condition (3.8) is very strong in the sense that it does not hold for Examples 3.3 and 3.4.

Theorem 3.2. (Prakasa Rao [70]) *Suppose that the hypotheses of Theorem* 3.1 *hold and that the noise* ε_t *are Gaussian. Then there exist positive constants* b_1 *and* b_2 *such that, for all sufficiently large* N *and all* $\varepsilon > 0$,

$$\mathbb{P}\big(\sqrt{N}(\hat{p}^N - p^*) > \varepsilon\big) < b_2 e^{-b_1 \varepsilon^2}.$$

If $\hat{p}^N$ is consistent then the normality of $\sqrt{N}(\hat{p}^N - p^*)$ is derived from the Taylor expansion and certain Central Limit Theorems; see Ivanov [39, 40].

In controlling dynamic systems, it is desirable that the states quickly settle and approach an equilibrium point at a high rate. This will result in die-off of the signal and, as a consequence, lack of observability. In such a case, condition (3.8) in Theorem 3.1 fails to hold and the estimator generally will be neither convergent nor consistent. We will investigate in the sequel whether it is possible to quantify the asymptotic behavior of a sequence of estimators in such cases and whether the sequence converges to an estimator that is unbiased and has a normal distribution.

We study, first, the behavior of the least squares function. Without loss of generality we assume that the errors $\varepsilon_t \sim \mathcal{N}(0, \mathbb{I}_m)$. Then,

$$\begin{aligned} L_N(y^N; p) = \frac{1}{2N} \sum_{t=1}^{N} (y_t - h_t(p^*))^T (y_t - h_t(p^*)) \\ - \frac{1}{N} \sum_{t=1}^{N} (y_t - h_t(p^*))^T (h_t(p) - h_t(p^*)) \\ + \frac{1}{2N} \sum_{i=1}^{N} (h_t(p) - h_t(p^*))^T (h_t(p) - h_t(p^*)). \end{aligned}$$

Substituting for y_t in terms of ε_t in these sums, we obtain

$$\begin{aligned} L_N(y^N; p) = \frac{1}{2N} \sum_{t=1}^{N} \varepsilon_t^T \varepsilon_t - \frac{1}{N} \sum_{t=1}^{N} (h_t(p) - h_t(p^*))^T \varepsilon_t \\ + \frac{1}{2N} \sum_{t=1}^{N} (h_t(p) - h_t(p^*))^T (h_t(p) - h_t(p^*)), \end{aligned} \tag{3.9}$$

and the first term in this expression represents the main contribution of the noise to the least squares function.

By Kolmogorov's Strong Law of Large Numbers (cf., Chung [18, Chapter 5]),

$$\frac{1}{N} \sum_{t=1}^{N} (h_t(p) - h_t(p^*))^T \varepsilon_t \to 0 \quad \text{a.s,} \tag{3.10}$$

and

$$\frac{1}{N} \sum_{t=1}^{N} \varepsilon_t^T \varepsilon_t \to m \quad \text{a.s.} \tag{3.11}$$

If we assume that $h_t(p)$ decays in the sense that $\max_{p\in\Theta}|h_t(p)| \to 0$ as $t \to \infty$, then the third term in (3.9) is of the form $N^{-1}\sum_{t=1}^{N} a_N$, with $a_N = (h_t(p) - h_t(p^*))^T(h_t(p) - h_t(p^*)$. By the Stolz–Cesàro Theorem [75, p. 80, Exercise 3.14], $N^{-1}\sum_{t=1}^{N} a_N \to 0$ a.s., so we obtain $\lim_{N\to\infty} L_N(y^N; p) = m$, a.s., hence the least squares function is asymptotically independent of p.

In this case, the problem of minimizing $L_N(y^N; p)$ with sufficiently large N is ill-posed in two senses. First, the solution to the least squares minimization problem is not unique. Second, in solving iteratively for p, the linearized approximation at the kth iteration has the form $A_k\hat{p}_k = b_k$ for some matrix A_k and some vector b_k; the least squares solution is $\hat{p}_k = (A_k^T A_k)^{-1}A_k^T b_k$, however the condition number of $A_k^T A_k$ is large ([79, Chapter 3], [83]). Therefore, from a computational point of view, we cannot expect that the least squares estimator will be consistent or will converge in distribution to a well-defined distribution.

Example 3.2. As a generalization of Example 3.1, consider the problem of least squares estimation for the regression model,

$$y_t = c_t p + \varepsilon_t, \quad t = 1, 2, \dots \tag{3.12}$$

where the regression coefficients c_t are constants and the errors ε_t are i.i.d with zero mean. Let $S_N^2 = \sum_{t=1}^{N} c_t^2$; then the least squares estimator of p based on measurements $y_1, \dots, y_N$ is $\hat{p}_{LS}^N = S_N^{-2}\sum_{t=1}^{N} c_t y_t$. Substituting for y_t in terms of ε_t we obtain

$$\hat{p}_{LS}^N - p = S_N^{-2}\sum_{t=1}^{N} c_t\varepsilon_t.$$

Suppose that ε_t are i.i.d with zero mean and variance σ^2, $\sigma > 0$. Then, $\mathbb{E}(\hat{p}_{LS}^N - p) = 0$, i.e., $\hat{p}_{LS}^N$ is an unbiased estimator of p. The variance of the error estimate is

$$\mathrm{Var}\,(\hat{p}_{LS}^N - p) = S_N^{-2}\sigma^2. \tag{3.13}$$

Consider the case in which $\lim_{N\to\infty} S_N^2 = \infty$. By (3.13), $\mathrm{Var}(\hat{p}_{LS}^N) \to 0$, so $\hat{p}_{LS}^N \to p$, a.s.; i.e., consistency holds. Furthermore, on applying the Lyapunov Central Limit Theorem, we obtain

$$S_N(\hat{p}_{LS}^N - p) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2).$$

Suppose now that $\lim_{N\to\infty} S_N = S < \infty$. In this case, since the variance of the error estimate (3.13) does not converge to 0, we lack consistency. Nevertheless, it can be shown (see [64]) that the limiting distribution of the random variable

$$Y_N = S_N^2(\hat{p}_{LS}^N - p) = \sum_{t=1}^{N} c_t\varepsilon_t \tag{3.14}$$

exists as $N \to \infty$, and we denote by Y the underlying limiting random variable.

Consider the case in which $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. Then $Y \sim \mathcal{N}(0, \sigma^2 S^2)$, so

$$S(\hat{p}_{LS}^N - p) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma^2).$$

Another case which has profound links to the theory of random walks is the situation in which ε_t is uniformly distributed on the set $\{-1, 1\}$, i.e., each ε_t is a unit random

walk. In this case, it follows from (3.14) that Y_N is the displacement from the origin of the random walker after N steps, where the step at time t is of length $|c_t|$. Since the characteristic functions of all $c_t\varepsilon_t$ are $\cos(c_t\xi)$ (see Chung [18]) and ε_t are independent, the characteristic function of Y_N is

$$\Phi_N(\xi) = \prod_{t=1}^{N} \cos(c_t\xi).$$

For particular values of c_t, we can obtain an explicit formula for $\lim_{N\to\infty} \Phi_N(\xi)$ as N tends to infinity; see [64]. In such cases, we will obtain the limiting distribution of Y_n. For example, if $c_t = 2^{-t}$ then $Y_n \to U(-1,1)$, the uniform distribution on $(-1,1)$; and if $c_t = 2^{(2-t)/2}$, then Y_n converges to a triangular distribution. In either of these two cases, Y has a symmetric distribution because ε_t does, but Y is not normal.

We now consider some examples of scalar models for which the measurement noise ε_t is assumed to be distributed as $\mathcal{N}(0,1)$.

Example 3.3. Consider the exponential decay regression model,

$$y_t = 10\exp(-tp) + \varepsilon_t, \quad t = 1, 2, \ldots,$$

where the unknown parameter $p \in \Theta$, a compact set in $\mathbb{R}^+$; this model was considered in [59] and [79, pp. 7-9].

The information matrix based on N measurements is

$$M_N(p) = 100 \sum_{t=1}^{N} t^2 \exp(-2tp).$$

Noting that

$$\begin{aligned} M_N(p) &= 25\left(\frac{\partial}{\partial p}\right)^2 \left(1 + \sum_{t=1}^{N} \exp(-2tp)\right) \\ &= 25\left(\frac{\partial}{\partial p}\right)^2 \frac{1-\exp(-2Np)}{1-\exp(-2p)}, \end{aligned}$$

where the sum is evaluated by virtue of being a geometric series, it follows that

$$\begin{aligned} \lim_{N\to\infty} M_N(p) &= 25\left(\frac{\partial}{\partial p}\right)^2 \frac{1}{1-\exp(-2p)} \\ &= 100\,\exp(-2p)(1+\exp(-2p))(1-\exp(2p))^{-3}; \end{aligned}$$

therefore, $\lim_{N\to\infty} M_N(p) < \infty$ for every $p \in \Theta$. Also, it follows from the Cramér-Rao inequality (3.5) that

$$\mathrm{Var}(\hat{p}^N) \geq \frac{1}{M_N(p)}$$

for any unbiased estimator $\hat{p}^N$ of p. Consequently, $\hat{p}^N$ does not converge to the true parameter p^*. On the other hand, since $\lim_{t\to\infty} \exp(-tp) = 0$ uniformly for $p \in \Theta$ then the least squares function is asymptotically independent of p.

We note that we can also apply the Bhattacharya inequality to functions of $\hat{p}^N$ to obtain similar results for estimating $g(p)$ for large classes of functions g.

Example 3.4. Consider the damped sinusoidal regression model,

$$y_t = \frac{10\sin(tp)}{t} + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{3.15}$$

where Θ is a closed interval of the form $[a, b]$, where a and b are given constants. We assume that $0 < a < b < \pi/2$, so that the function $\sin p$ is injective on Θ. The model (3.15) arises in the regularization of inverse problems [24, pp. 4–6].

For this example, the information matrix is

$$M_N(p) = 100 \sum_{t=1}^{N} \cos^2(tp).$$

Applying the trigonometric identity, $\cos^2(tp) = \frac{1}{2}(1 + \cos(2tp))$, and the well-known formula,

$$\sum_{t=1}^{N} \cos(t\alpha) = \frac{\sin((2N+1)\alpha/2) - \sin(\alpha/2)}{2\sin(\alpha/2)}, \tag{3.16}$$

we obtain

$$M_N(p) = 50 \sum_{t=1}^{N} (1 + \cos(2tp)) = 50\Big[N + \frac{\sin((2N+1)p)}{2\sin(p)} - \frac{1}{2}\Big].$$

Therefore $M_N(p) \approx 50N$, and hence $M_N(p) \to \infty$, as $N \to \infty$. Note that Lemma 3.6 provides no information about this case because the condition that $M_N(p) \to \infty$ as $N \to \infty$ is only a necessary condition for consistency. Moreover, since $10\sin(tp)/t \to 0$ as $t \to \infty$, this example shares with the previous example the issue that the least squares function is asymptotically independent of p.

The next example is more subtle, and it requires a correspondingly more detailed analysis. Kundu [51] treated a similar example and gave a proof of the strong consistency of the least squares estimator, and we will obtain an alternative proof of that result.

Example 3.5. Consider the undamped sinusoidal regression model,

$$y_t = 10\sin(tp) + \varepsilon_t, \quad t = 1, 2, \ldots$$

where $p \in \Theta = [a, b]$ as in Example 3.4. In this model, the signal does not decay; i.e, $h(t; p) = 10\sin(tp)$ does not converge to 0 as $t \to \infty$. Also, the corresponding information matrix is

$$M_N(p) = 100 \sum_{t=1}^{N} t^2 \cos^2(tp).$$

By comparison with the information matrix in Example 3.4, we find that $M_N(p) \to \infty$ as $N \to \infty$.

Applying (3.9), (3.10), and (3.11), we find that the least squares function satisfies

$$\frac{L_N(y^N; p)}{100} \approx \frac{1}{2} + \frac{1}{2N} \sum_{t=1}^{N} (\cos(tp) - \cos(tp^*))^2.$$

Expanding the squared term in the sum and changing the squared cosine terms to the cosine of double angles, we obtain

$$\begin{aligned}\frac{L_N(y^N;p)}{100} &\approx \frac{1}{2}+\frac{1}{2N}\sum_{t=1}^{N}\Big[\frac{1+\cos(2tp)}{2}+\frac{1+\cos(2tp^*)}{2}\\ &\qquad -\cos(t(p+p^*))-\cos(t(p-p^*))\Big]\\ &= 1+\frac{1}{4N}\sum_{t=1}^{N}\big(\cos(2tp)+\cos(2tp^*)\big)\\ &\quad -\frac{1}{2N}\sum_{t=1}^{N}\big[\cos(t(p+p^*))-\cos(t(p-p^*))\big]\end{aligned}$$

Applying the identity (3.16) we deduce that, for fixed $p \neq p^*$,

$$\begin{aligned}\frac{L_N(y^N;p)}{100} \approx 1 &+ \frac{\sin((2N+1)tp)-\sin(p)}{8N\sin(p)}+\frac{\sin((2N+1)tp^*)-\sin(p^*)}{8N\sin(p^*)}\\ &- \frac{\sin((2N+1)t(p+p^*)/2)-\sin((p+p^*)/2)}{4N\sin((p+p^*)/2)}\\ &- \frac{\sin((2N+1)t(p-p^*)/2)-\sin((p-p^*)/2)}{4N\sin((p-p^*)/2)}.\end{aligned}$$

Hence, $\lim_{N\to\infty} L_N(y^N;p) = 100$, a.s. However, $L_N(y^N,p^*) = 50$ for all finite N; therefore, the limit function $\bar{L}(p) = \lim_{N\to\infty} L_N(y^N;p)$ exists but is not continuous. In particular, the results in [43] which require the continuity of $\bar{L}$ are not applicable.

To establish the consistency of the least squares estimator for this model, we start with the observation that p^* is the unique minimizer of $\bar{L}(p)$. Letting $(\Omega,\mathbb{P})$ be the probability space on which Y_k are defined, we set

$$B=\big\{\omega\in\Omega\,|\,\forall\,\varepsilon>0,\exists\,N_0>0,\forall\,N>N_0,\ \big|L_N(y^N;p)-\bar{L}(p)\big|<\varepsilon\quad\forall\,p\in\Theta\big\}. \tag{3.17}$$

Since $L_N(y^N;p)\to\bar{L}(p)$, a.s., uniformly with respect to p, we have $\mathbb{P}(B)=1$. For any $\omega\in B$, let $\varepsilon=1$ and choose a corresponding N_0 in (3.17). For $N>N_0$, if $\hat{p}^N\neq p^*$, then $L_N(y^N;\hat{p}^N)=100$. However, $L_N(y^N;p^*)=50$, which contradicts the minimality of $\hat{p}^N(\omega)$. Therefore, $\hat{p}^N(\omega)=p^*$, i.e., as $n\to\infty$, $\hat{p}^N(\omega)\to p^*$ for all $\omega\in B$. Because $\mathbb{P}(B)=1$ then it follows that $\hat{p}^N$ converges to p^* almost surely.

Remark 3.1. We proved earlier the consistency of the least squares estimators for Example 3.2. However, as the sample size N increases, the least squares function becomes ill-posed and it may have many local optima, causing difficulties in solving the estimation problem numerically. These issues will be illustrated in Section 3.4.

3.4 Numerical considerations

3.4.1 Sequential least squares strategy

To establish the asymptotic properties of the least squares estimator, it is assumed that the least squares minimization problem can be solved exactly for each random sample, i.e., the solution obtained is the global minimizer of the least squares function.

However, one difficulty here is that numerical software often finds local solutions to the minimization problem. On the other hand, in Examples 3.3 and 3.4, the least squares functions for large samples are nearly constant, so the parameter estimation problem can become ill-posed and the numerical algorithms can terminate far from the desired solution. However, by implementing the *sequential LS*, we can obtain convergence of the resulting estimates to the true parameter values as the sample size increases.

The Sequential LS: Choose $k > 0$, the increment number of new measurements.

- Step 1: Start from N_0 measurements, at least equal to the number of parameters. Compute the estimate $\hat{p}^{N_0}$ and an estimate of its covariance C_{N_0}.
- Step 2: Collect k new measurements. Solve the regularized least squares problem,

$$\min_p \Big[L_{N_0+k}(y^{N_0+k}; p) + \frac{1}{2}(p - \hat{p}^{N_0})^T C_{N_0}^{-1}(p - \hat{p}^{N_0})\Big],$$

  to obtain $\hat{p}^{N_0+k}$ and C_{N_0+k}.
- Step 3: Set $N_0 \leftarrow N_0 + k$. Stop if no new measurements are available. Otherwise, return to step 2.

Consider numerical algorithms to solve the parameter estimation problem, for example the Gauss-Newton method. Near the least squares solution $\hat{p}^N$, by linearization, the iterations are approximately normally distributed around $\hat{p}^N$; see Bock [11], Körkel [46], Schlöder [78]. As the sample size is small, estimates may already be close to the true value p^*. When $\hat{p}^N$ is already inside a suitably small region around p^*, adding new measurements will help bring the estimates closer and closer to the true value. As a result, the estimates will converge to a well-defined distribution.

If the information matrix tends to infinity, the variance decreases to 0; so, as in Examples 3.4 and 3.5, we have consistency. The Extended Kalman Filter (EKF) and its variants may also be applicable, see [57, 80], but they have the usual drawback of EKF that they may not converge for strongly nonlinear systems (see [33]). If one does not use sequential least squares together with some regularization then, for large N and depending on the choice of the initial iteration point, $\hat{p}^N$ can be far from the true parameter. Rigorous proofs of these results will be an objective of future research. As a first step toward those results, we provide in Section 3.5 an algorithm to implement the sequential LS strategy and we prove a local convergence result for the algorithm.

3.4.2 Numerical examples

The simulation below is done in MATLAB with `lsqnonlin` with the *trust region reflective* method as a solver for the least squares problem. We consider Example 3.4. Take the true value $p^* = 1.0$ and the initial guess $p_0 = 0.5$. We run 500 estimation problems, each with $N = 100$ measurements at $t = 1, \ldots, 100$. Measurements are generated by adding standard white noise to $h_t(p^*)$ with the function `randn`. By the traditional LS, we mean that the least squares function is formed by all N measurements. We estimate the parameter p by the traditional least squares and sequential LS. The means and standard deviations of the estimates are then computed. The results of simulations for Example 3.4 are displayed in Figures 3.1 and 3.2. For the sequential least squares with increment $k = 5$, we have

$$\hat{p}_s = 0.9949 \pm 0.0969,$$

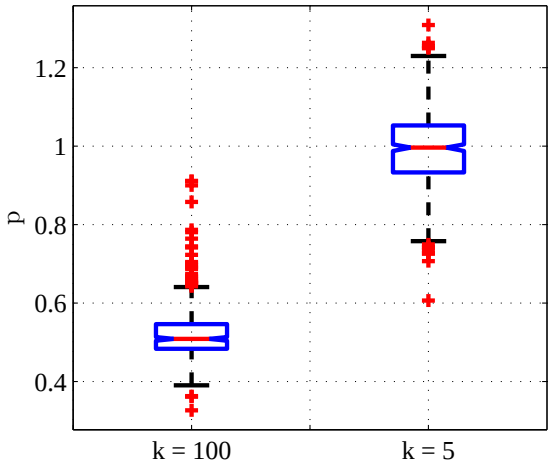

Figure 3.1: Boxplots of least squares estimates for Example 3.4. The traditional least squares procedure ($k = 100$) leads to estimates trapped around the initial guess and failing to converge to the true value. Estimates produced by the sequential least squares method ($k = 5$) show a trend of convergence to the true value of the parameter and with a well-defined distribution.

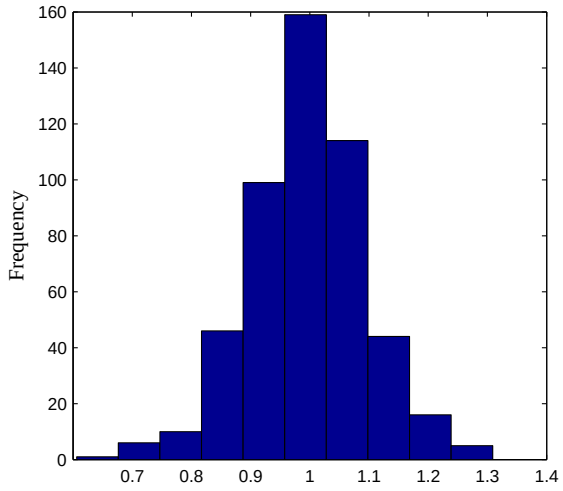

Figure 3.2: This diagram provides a histogram of estimates derived by the sequential least squares method. The histogram exhibits the normality of the estimates.

while for the traditional LS,

$$\hat{p} = 0.5239 \pm 0.0696.$$

In the boxplots, the central box spans the quartiles 25%, 50%, 75% consecutively upward. The line in the box marks the median. Data more than 1.5 times of box length from the median are plotted individually as possible outliers, marked by + signs. See Moore and McCabe [63] for more detail on boxplots.

The normality of the estimates is evident from Figure 3.2, and a χ^2-square goodness-of-fit test for normality provided further confirmation of the normal distribution of the estimates. Moreover, similar results were obtained for Example 3.5.

Remark 3.2. Gauss and the sequential LS
The idea of sequential least squares seems to be natural in practice, especially when measurements are recorded in a long period. Instead of waiting to have a full set of measurements, we use available measurements to get rough approximations. Additional new data will improve our estimates.

We remark that the sequential least squares method can be traced back to Gauss, when he studied astronomy more than 200 years ago. Gauss, in his work *Theoria motus corporum coelestium in sectionibusconicis Solem ambientium*, described his two-step method for determining the orbit of a celestial body, Ceres, from actual observations. The first was to obtain an approximate solution using an initial, small number of observations; and the second step was to improve the result from the first step with the help of additional observational data. We refer to Bühler [15, Chapter 8] for a comprehensive account of the history of Gauss' calculation of the orbit of Ceres.

3.5 Convergence of a Gauss-Newton method for sequential least squares problems

In this section, we provide an algorithm for applying the sequential least squares strategy, and we prove a local convergence result for the algorithm.

Consider the problem of estimating the parameter p in the model,

$$y_t = h_t(p) + \varepsilon_t, \qquad t = 1, 2, ... \tag{3.18}$$

where $p \in \Theta$, a compact subset of $\mathbb{R}^{n_p}$; $h_t : \Theta \to \mathbb{R}$ are measurement functions, $t = 1, \ldots, N$; $y_1, \ldots, y_N \in \mathbb{R}$ is a sample of noisy measurements with independent identically distributed Gaussian noise $\varepsilon_t \sim \mathcal{N}(0, 1)$, $t = 1, \ldots, N$. Then the well-known least squares method for parameter estimation is to calculate

$$\min_{p \in \Theta} L_N(p), \tag{3.19}$$

the minimum of the least squares function

$$L_N(p) = \frac{1}{2N} \sum_{t=1}^{N} \big(y_t - h_t(p)\big)^2.$$

We shall make the following assumptions:

(H1) There exists for each N a global minimizer, $p_N^* \in \Theta$, of (3.19).

(H2) The sequence $L_N(p)$ converges uniformly on Θ to a function $L(p)$.

Because of the uniform convergence, (H2) implies that $L(p)$ is continuous on Θ.

Proposition 3.1. *Suppose that $\bar{p} \in \Theta$ is an accumulation point of the sequence (p_N^*). Then $\bar{p}$ is a minimizer of the function $L(\cdot)$ on Θ. Furthermore, if $L(\cdot)$ has a unique minimizer p^* then $\lim_{N\to\infty} p_N^* = p^*$.*

Proof. Since $\bar{p}$ is an accumulation point of the sequence (p_N^*) then by passing to a subsequence we may assume, without loss of generality, that

$$\lim_{N \to \infty} p_N^* = \bar{p}. \tag{3.20}$$

Moreover, since Θ is compact and $L(\cdot)$ is continuous on Θ, there exists $\tilde{p} \in \Theta$ such that $L(\tilde{p}) = \min_{p\in\Theta} L(p)$.

We need to show that $L(\bar{p}) = L(\tilde{p})$, and we proceed by contradiction. Suppose that $L(\bar{p}) > L(\tilde{p})$, and set $\delta = (L(\bar{p}) - L(\tilde{p}))/4$. Then there exists a neighborhood U of $\bar{p}$ such that $L(p) > L(\bar{p}) - \delta$ for all $p \in U$. Since (3.20) holds then there exists N_0 such that $p_N^* \in U$ for all $N \geq N_0$.

By the uniform convergence of L_N to L, there exists $N_1 > N_0$ such that $|L_N(p) - L(p)| < \delta$ for all $p \in U$ and $N \geq N_1$. Consequently, by the definition of δ, we have

$$L_{N_1}(\tilde{p}) < L(\tilde{p}) + \delta < L(\bar{p}) - 3\delta < L(p_{N_1}^*) - 2\delta < L_{N_1}(p_{N_1}^*) - \delta.$$

However, this contradicts the assumption that $p_{N_1}^*$ is a minimizer of L_{N_1} on Θ. Hence the first part of the proposition is proved.

For the second part we suppose, to the contrary, that $p_N^* \not\to p^*$ as $N \to \infty$. Then there exists $\delta_1 > 0$ and a subsequence $(p_{N_m}^*)$ of (p_N^*) such that $\|p_{N_m}^* - p^*\| > \delta_1$ for all N_m. Since Θ is compact and $(p_{N_m}^*) \subset \Theta$, there exists a subsequence, also denoted by $(p_{N_m}^*)$ which converges to a minimizer of $L(p)$. Since p^* is the unique minimizer of $L(p)$, then it must hold that $\lim_{N_m\to\infty} p_{N_m}^* = p^*$. This results in a contradiction and hence completes the proof. □

In view of Proposition 3.1, we now assume that

(H3) $L(p)$ possesses a unique minimizer p^* in Θ.

We remark that the hypotheses (H1)-(H3) are basic in the literature on the theoretical convergence properties of the least squares, see e.g., [39, 40, 43]. We now devise an iterative procedure, the Gauss-Newton method, which converges to p^* in this case. We set

$$\begin{aligned} F_N(p) &= \big(y_t - h_t(p)\big)_{t=1}^N \in \mathbb{R}^N, \\ J_N(p) &= \frac{\partial}{\partial p} F_N(p) \in \mathbb{R}^{N\times p}, \end{aligned} \tag{3.21}$$

and

$$M_N(p) = \frac{1}{N} J_N^T(p) J_N(p). \tag{3.22}$$

Because $L_N(p)$ attains its minimum at p_N^*, there holds

$$\frac{1}{N} J_N^T(p_N^*) F_N(p_N^*) = \frac{\partial}{\partial p} L_N(p)\Big|_{p=p_N^*} = 0.$$

We assume that there exist $N_0 \geq p$ and a neighborhood U of p^* such that $J_N(p)$ has full rank for all $p \in U$ and $N \geq N_0$. Consequently, $M_N(p)$ is nonsingular for all $p \in U$ and $N \geq N_0$.

We now choose a small tolerance, `tol` > 0 and consider the following scheme for approximating p^*:

1. Starting at $N = N_0$, we initiate the iterative procedure at p_{N_0}.

2. The iterative step:

$$p_{N+1} = p_N - \frac{1}{N} M_N^{-1}(p_N) J_N^T(p_N) F_N(p_N). \tag{3.23}$$

3. If $||p_{N+1} - p_N|| < \texttt{tol}$, stop. Else set $N \to N+1$ and return to 2.

In the following we will show that if p_{N_0} is sufficiently close to p^* then $\lim_{N\to\infty} p_N = p^*$. For this purpose, we impose the following assumptions, cf. Bock [11]:

(H4) There exists $\beta \in [0,1)$ such that, for all $N \geq N_0$ and $p \in \Theta$,

$$\left\| \frac{1}{N} M_N^{-1}(p) \left(J_N^T(p) - J_N^T(p_N^*) \right) F_N(p_N^*) \right\| \leq \beta \left\| p - p_N^* \right\|.$$

(H5) There exists $\omega \in (0,\infty)$ such that for all $N \geq N_0$, $p \in \Theta$, and $0 \leq t \leq 1$,

$$\frac{1}{N} \left\| M_N^{-1}(p) J_N^T(p) \left(J_N^T(p) - J_N(p_N^* + t(p - p_N^*)) \right) (p - p_N^*) \right\| \leq \omega t \left\| p - p_N^* \right\|^2.$$

We now prove the following result.

Theorem 3.3. *Suppose that* (H1)–(H5) *hold. If the initial iteration point p_{N_0} is chosen such that*

$$\omega \left\| p_{N_0} - p_{N_0}^* \right\| + \beta < 1$$

and

$$\left\| p_{N+1}^* - p_N^* \right\| < \frac{(1-\beta)^2}{2\omega} \text{ for all } N \geq N_0,$$

then $p_N \to p^*$ as $N \to \infty$.

Proof. By (3.23), we have

$$\begin{aligned} p_{N+1} - p_N^* &= p_N - p_N^* - \frac{1}{N} M_N^{-1}(p_N) J_N^T(p_N) F_N(p_N) \\ &= M_N^{-1}(p_N) \Big\{ M_N(p_N)(p_N - p_N^*) \\ &\qquad - \frac{1}{N} \big(J_N^T(p_N) F_N(p_N) - J_N^T(p_N^*) F_N(p_N^*) \big) \Big\}. \end{aligned}$$

Substituting for the inner $M_N(p_N)$ term using (3.22), and adding and subtracting a term inside the braces, we obtain

$$\begin{aligned} p_{N+1} - p_N^* &= \frac{1}{N} M_N^{-1}(p_N) \Big\{ - \big(J_N^T(p_N) - J_N^T(p_N^*) \big) F_N(p_N^*) \\ &\qquad + J_N^T(p_N) J_N(p_N)(p_N - p_N^*) - J_N^T(p_N) \big(F_N(p_N) - F_N(p_N^*) \big) \Big\} \\ &= -\frac{1}{N} M_N^{-1}(p_N) \big(J_N^T(p_N) - J_N^T(p_N^*) \big) F_N(p_N^*) \\ &\quad + \frac{1}{N} M_N^{-1}(p_N) J_N^T(p_N) \big[J_N(p_N)(p_N - p_N^*) - \big(F_N(p_N) - F_N(p_N^*) \big) \big]. \end{aligned} \tag{3.24}$$

By (3.21),

$$F_N(p_N) - F_N(p_N^*) = \int_0^1 J_N(p_N^* + t(p_N - p_N^*))(p_N - p_N^*) dt;$$

substituting this expression into (3.24) we obtain

$$p_{N+1} - p_N^* = -\frac{1}{N} M_N^{-1}(p_N)(J_N^T(p_N) - J_N^T(p_N^*)) F_N(p_N^*)$$

$$+\int_0^1 \frac{1}{N} M_N^{-1}(p_N) J_N^T(p_N) \Big\{ J_N(p_N) - J_N(p_N^* + t(p_N - p_N^*)) \Big\} (p_N - p_N^*) dt.$$

Now we apply the hypotheses (H4)-(H5) to derive the inequalities

$$\begin{aligned} \|p_{N+1} - p_N^*\| &\le \beta \|p_N - p_N^*\| + \int_0^1 \omega t \|p_N - p_N^*\|^2 dt \\ &\le \beta \|p_N - p_N^*\| + \frac{\omega}{2} \|p_N - p_N^*\|^2 . \end{aligned}$$

It follows that

$$\left\|p_{N+1} - p_{N+1}^*\right\| \le \beta \|p_N - p_N^*\| + \frac{\omega}{2} \|p_N - p_N^*\|^2 + \left\|p_{N+1}^* - p_N^*\right\| . \tag{3.25}$$

Set $v_N = \|p_N - p_N^*\|$, $w_N = \left\|p_{N+1}^* - p_N^*\right\|$ and $\alpha = \omega/2$. From the hypotheses, we deduce that $w_N < (1-\beta)^2/4\alpha$ for all $N \ge N_0$ and $v_{N_0} < (1-\beta)/2\alpha$. Applying Lemma 3.7 below, we deduce that $v_N \to 0$ as $N \to \infty$. By Proposition 3.1, $p_N^* \to p^*$, therefore, $p_N \to p^*$ as $N \to \infty$. □

Lemma 3.7. *Suppose that nonnegative sequences* (v_n) *and* (w_n), $n = 0, 1, 2, \ldots$, *satisfy* $\lim_{n\to\infty} w_n = 0$ *and*

$$v_{n+1} \le \alpha v_n^2 + \beta v_n + w_n \tag{3.26}$$

for constants $\alpha > 0$ *and* $0 \le \beta < 1$. *If* $w_n < (1-\beta)^2/4\alpha$ *and* $v_0 < (1-\beta)/2\alpha$ *for all* n, *then* $\lim_{n\to\infty} v_n = 0$.

Proof. By (3.26), we have

$$v_1 \le \alpha \left(\frac{1-\beta}{2\alpha}\right)^2 + \beta \left(\frac{1-\beta}{2\alpha}\right) + \frac{(1-\beta)^2}{4\alpha} = \frac{1-\beta}{2\alpha}.$$

Continuing this way, it is straightforward to prove by induction that $v_n < (1-\beta)/2\alpha$ for all $n \ge 0$, i.e., $\alpha v_n + \beta < (1+\beta)/2$. As a consequence, it follows from (3.26) that

$$v_{n+1} \le (\alpha v_n + \beta) v_n + w_n \le \gamma v_n + w_n \tag{3.27}$$

for all $n \ge 0$, where $\gamma = (1+\beta)/2 \in [1/2, 1)$.

For $\delta > 0$, choose N_0 so that $0 \le w_n < \delta(1-\gamma)/2$ for all $n \ge N_0$. Also choose N_1 such that $\gamma^n < \delta\alpha/(1-\beta)$ for all $n \ge N_1$. By repeatedly applying (3.27), we obtain for $n > N_0 + N_1$,

$$\begin{aligned} v_n &\le \gamma^{n-N_0} v_{N_0} + \gamma^{n-N_0-1} w_{N_0} + \gamma^{n-N_0-2} w_{N_0+1} + \cdots + \gamma^0 w_{n-1} \\ &\le \frac{\delta\alpha}{1-\beta} \frac{1-\beta}{2\alpha} + \frac{\delta}{2}(1-\gamma)(\gamma^{n-N_0-1} + \cdots + \gamma + 1) \\ &= \frac{\delta}{2} + \frac{\delta}{2}(1-\gamma)\frac{1-\gamma^{n-N_0}}{1-\gamma} < \delta. \end{aligned}$$

Therefore, $v_n \to 0$ as $n \to \infty$. □

Chapter 4

Finite Support for Optimal Experimental Designs

In the continuous case, Optimal Experimental Design (OED) deals with designs that are described by probability distributions or samples over the experimental domain. An optimal design may correspond to a distribution having finite or infinite support or being continuous. In this chapter, the structure of optimal samples for experimental designs is elucidated. It is shown that any design is in fact equivalent to a design with a finite number of support points. The lower bound and upper bound of this number, especially for optimal designs, are given and examples indicate their sharpness. Moreover, we propose an algorithm to construct optimal designs which have finite support. Several applications to OED for dynamic systems with inputs are also discussed.

4.1 Introduction

Consider the dynamic process modeled by an initial value problem (IVP) with inputs

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t), p), \quad t \in [t_0, t_f], \\ x(t_0) = x_0(p), \end{cases} \tag{4.1}$$

where $x(t) \in \mathbb{R}^{n_x}$ are the states, $u(t) \in \mathbb{U} \subseteq \mathbb{R}^{n_u}$ are the inputs and $p \in \mathbb{R}^k$ are unknown parameters. In order to get data to estimate p, we consider a measurement function $\hat{\eta}(\cdot)$. The measurements $\hat{y}(t) \in \mathbb{R}^{n_y}$ are usually corrupted by noise $\hat{\varepsilon}(t)$

$$\hat{y}(t) = \hat{\eta}(x(t; t_0, x_0(p), u(\cdot), p), p) + \hat{\varepsilon}(t).$$

The quality of estimates is often expressed in terms of the covariance matrix of the estimated parameters. Optimal Experimental Design (OED) aims to minimize some optimality criterion acting on this covariance matrix by appropriately choosing the experimental conditions such as the inputs u and the time points t at which $\hat{y}$ is evaluated.

The widely studied case in the literature is discrete designs where the set of competing designs consists of discrete probability distributions over a discrete experimental domain. In spite of this convention, it is still desirable to have freedom in choosing sampling points. We thus enlarge the set of competing designs to all general probability distributions which are called continuous designs. The following questions arise naturally. Firstly, do there exist better designs in this case compared with the discrete case?

Secondly, what is the most parsimonious way of describing the optimal designs? These problems were occasionally mentioned in the literature, starting in Pukelsheim [69], Pázman [66]. They considered the discrete case and discussed the question of minimal number of support points for optimal designs. Our objective is to give a comprehensive investigation of this issue. We treat the problem for the continuous case in full generality using tools from Functional Analysis and Convex Analysis. It is then shown that any design is equivalent to a discrete design with finitely many support points. The upper bound and lower bound for the support size for the optimal designs are established. Obtained results are illustrated by examples, involving also dynamic processes.

The chapter is organized as follows. Section 4.2 formulates the problem of continuous OED. In Sections 4.3-4.4, we show that any continuous designs can be reduced to discrete designs with finite support and give the bounds on the size of support for optimal designs. Applications for initial value problems (IVPs) with inputs are delivered in Section 4.5. Section 4.6 proposes a problem setting to construct the discrete optimal designs. The chapter concludes with examples on the sharpness of the bounds given in Section 4.4 and OED for dynamic systems.

4.2 Formulation of the continuous optimal experimental design problem

In the following, we consider a generic nonlinear model

$$y(q) = \eta(q,p) + \varepsilon(q), \tag{4.2}$$

with $q \in Q$, where Q is a compact set in $\mathbb{R}$ (or $\mathbb{R}^m$) called *experimental domain*; $p \in \mathbb{R}^k$ are unknown parameters; $\eta(\cdot)$ is the measurement function; $y(q) \in \mathbb{R}^{n_y}$ are noisy measurements, and $\varepsilon(q) \in \mathbb{R}^{n_y}$ are random errors satisfying [1]

$$\mathbb{E}(\varepsilon(q)) = 0; \quad \mathbb{E}(\varepsilon(q_1)\varepsilon(q_2)) = \delta_{q_1 q_2}\mathbb{I}_{n_y}. \tag{4.3}$$

where $\delta_{q_1 q_2} = 1$ if $q_1 = q_2$ and 0 otherwise; $\mathbb{I}_{n_y}$ is the identity matrix of size n_y.

Remark 4.1. *In connection with* IVP (4.1)*, suppose that we use a discretization of the inputs u on the grid $t_0 < t_1 < t_2 < ... < t_N = t_f$, e.g., $u(t) = u_i \in \mathbb{R}^{n_u}$, $t_i \le t < t_{i+1}$ for $i = 0, 1, ..., N-1$. We can write $q = (t, u_0, u_1, ..., u_{N-1}) \in \mathbb{R}^{Nn_u+1}$ and consider $\hat{y}, \hat{\eta}$ and $\hat{\varepsilon}$ as functions of q. Furthermore we could consider $\hat{\eta}$ as a function of (t,p) and write $t = q$. By those ways we arrive at special cases of model* (4.2)*.*

By an (experimental) design on Q, we understand a probability distribution ξ on Q.

We make the assumption that η is differentiable with respect to p and z is continuous in q, where we denote

$$z(q) = \frac{\partial \eta}{\partial p}(q,p); \quad g(q) = z(q)^T z(q).$$

Note that $z(\cdot)$ and $g(\cdot)$ depend on p, but since p is fixed for considered designs, we omit p in their arguments for brevity. Now suppose that ξ is a Borel measure. This ensures

[1] It should be $\mathbb{E}(\varepsilon(q_1)\varepsilon(q_2)) = \delta_{q_1 q_2}\Sigma^2(q_1)$, in which $\Sigma(q_1)$ is a diagonal matrix representing the standard deviation of noise in the measurements. However, by suitable scaling, we can assume without loss of generality that $\Sigma(q_1) = \mathbb{I}_{n_y}$.

the existence of the following integral

$$M(\xi) = M(\xi, p) = \int_Q \left(\frac{\partial \eta}{\partial p}(q,p)\right)^T \left(\frac{\partial \eta}{\partial p}(q,p)\right) d\xi(q). \tag{4.4}$$

For a fixed p, the *information* matrix corresponding to a design ξ is defined by $M(\xi, p)$. It is easy to verify that $M(\xi)$ is symmetric, positive semidefinite. So we can consider $M(\xi)$ as elements of the Euclidean space $X = \mathbb{R}^{k(k+1)/2}$. Furthermore, we recall the following property of integrals, see Rudin [76]: For an arbitrary probability measure μ on Q and any continuous linear functional Λ in X, it holds that

$$\Lambda\left(\int_Q g(q)d\mu\right) = \int_Q \Lambda g(q) d\mu(q), \tag{4.5}$$

A *finite support design* on Q is a discrete distribution ξ on Q such that

$$\text{supp}(\xi) = \{q \in Q \mid \xi(q) > 0\} \text{ consists of finitely many points } q_i; \sum_{q_i \in \text{supp}(\xi)} \xi(q_i) = 1.$$

In this case, $\xi(q_i)$ is called the *weight* at q_i. The information matrix then can be written as

$$M(\xi) = M(\xi, p) = \sum_{q_i \in \text{supp}(\xi)} \xi(q_i) \left(\frac{\partial \eta}{\partial p}(q_i,p)\right)^T \left(\frac{\partial \eta}{\partial p}(q_i,p)\right).$$

The inverse of M, in case it exists,

$$C(\xi) = C(\xi, p) = M^{-1}(\xi, p),$$

is called *variance-covariance* or simply *covariance* matrix.

To assess the quality of designs, we define scalar functions $\mathscr{K}(C)$ acting on the set of possible covariance matrices. Some well-known ones are

$$\mathscr{K}_D(C) = (\det(C))^{\frac{1}{k}} \quad \text{(D-criterion)}; \quad \mathscr{K}_A(C) = \frac{1}{k}\text{Trace}(C) \quad \text{(A-criterion)}.$$

OED for a *fixed* p aims to minimize one of these functionals over all designs ξ of interest.

Remark 4.2. *In case M is singular with a specific choice of ξ, we simply set $\mathscr{K} = \infty$ without changing the result of the optimization problem. From now on, the notation $\mathscr{K}$ is generically used for $\mathscr{K}_D$ or $\mathscr{K}_A$.*

Several questions arise. How many measurements do we need at least to identify p? Can we achieve (optimal) designs by discrete designs, especially designs with finite support? And if so, how can we choose support points to construct optimal designs? Those will be treated in the following sections.

4.3 Reduction to finite support designs

For a set $E \subseteq \mathbb{R}^n$, the *convex hull* of E, denoted by $\text{conv}(E)$, is the set of all convex combinations of the points in E, i.e.,

$$\text{conv}(E) = \left\{ x \in \mathbb{R}^n \;\middle|\; \exists m \geq 1,\ x_i \in E,\ \lambda_i \geq 0, i = 1, ..., m,\ \sum_{i=1}^m \lambda_i = 1,\ \ x = \sum_{i=1}^m \lambda_i x_i \right\}.$$

To make our exposition self-contained, we present here some standard results of Convex Analysis which can be found in Rockafellar and Wets [74], Rudin [76].

Lemma 4.1. (Rudin [76, p. 72]) *If $E \subseteq \mathbb{R}^n, x \in \text{conv}(E)$, then x lies in the convex hull of some subset of E which contains at most $n+1$ points.*

Proof. We will show that if $k > n$ and x is represented by $k+1$ points in E, then only k points are needed. The desired conclusion follows by induction. Suppose that

$$x = \sum_{i=1}^{k+1} \lambda_i x_i, x_i \in E,$$

where $\lambda_i > 0$, $\sum_{i=1}^{k+1} \lambda_i = 1$. Consider the linear mapping $L : \mathbb{R}^{k+1} \longrightarrow \mathbb{R}^{n+1}$ defined by

$$L(a_1, a_2, ..., a_{k+1}) = \left(\sum_{i=1}^{k+1} a_i x_i, \quad \sum_{i=1}^{k+1} a_i \right).$$

Since $k > n$, the null-space of L must be nonzero. Hence there exist $a_1, ..., a_{k+1}$, at least one of which is greater than 0 such that

$$\sum_{i=1}^{k+1} a_i x_i = 0; \quad \sum_{i=1}^{k+1} a_i = 0.$$

Set

$$t = \min_i \{\lambda_i / a_i, \ a_i > 0\} > 0.$$

Then $\lambda_i - t a_i \geq 0$ for all i and $\lambda_i - t a_i = 0$ for at least one i. Moreover

$$x = \sum_{i=1}^{k+1} \lambda_i x_i - t \sum_{i=1}^{k+1} a_i x_i = \sum_{i=1}^{k+1} (\lambda_i - t a_i) x_i.$$

This completes the proof. □

Lemma 4.2. (Rudin [76, p. 73]) *If K is a compact set in $\mathbb{R}^n$ then so is* conv(K).

Proof. Let S be the unit simplex in $\mathbb{R}^{n+1}$, i.e.,

$$S = \left\{ \lambda \in \mathbb{R}^{n+1}, \quad \sum_{i=1}^{n+1} \lambda_i = 1, \ \lambda_i \geq 0 \right\}.$$

It follows from Lemma 4.1 that $x \in \text{conv}(K)$ if and only if there are $x_i \in K$, $i = 1, 2, ..., n+1$ and $\lambda = (\lambda_1, \lambda_2, ..., \lambda_{n+1}) \in S$ such that

$$x = \lambda_1 x_1 + \lambda_2 x_2 + ... + \lambda_{n+1} x_{n+1}.$$

Define the mapping $L : S \times K^{n+1} \longrightarrow \mathbb{R}^n$ by

$$L(\lambda, x_1, x_2, ..., x_{n+1}) = \lambda_1 x_1 + \lambda_2 x_2 + ... + \lambda_{n+1} x_{n+1}.$$

Obviously, L is continuous and $\text{conv}(K) = L(S \times K^{n+1})$. Therefore, conv$(K)$ is compact, since both S and K are compact.

□

Lemma 4.3. (A separation theorem, Rockafellar and Wets [74, p. 63]) *Let K be a closed convex set in $\mathbb{R}^n$, $u \notin K$. Then there exist $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ such that*

$$a^T u > \alpha > a^T x, \ \forall x \in K.$$

Proof. It is well-known that there exists a unique $z \in K$ such that

$$\|u - z\| = d(u, K) = \inf\{\|u - x\|, \ x \in K\} > 0,$$

and

$$(u - z, \ x - z) \leq 0, \text{ for all } x \in K.$$

($\|\cdot\|$ is the Euclidean norm, $(\cdot, \cdot)$ denotes the scalar product). We have

$$0 \geq (u - z, \ x - z) = (u - z, \ u - z) - (u - z, \ u - x).$$

Set $a = u - z$. If follows that

$$a^T u - a^T x \geq \|u - z\|^2 > \|u - z\|^2/2 > 0.$$

The desired conclusion follows after setting $\alpha = a^T u - \frac{\|u-z\|^2}{2}$.

□

We are now in the position to state the core result of this section. Recall that k is the dimension of the parameters p and $g(q) : Q \longrightarrow X$, $X = \mathbb{R}^{k(k+1)/2}$ is a continuous mapping. Since Q is compact, $g(Q)$ is compact. By Lemma 4.2, its convex hull $H = \text{conv}(g(Q))$ is also compact.

Theorem 4.1. *The value of the integral $\int_Q g(q)d\mu$ lies in H. As a result, any information matrix can be constructed from at most $\frac{k(k+1)}{2} + 1$ points $q \in Q$.*

Proof. Set $m = \int_Q g(q)d\mu(q)$. Suppose on the contrary that $m \notin H$. Since H is closed and convex, there exist by Lemma 4.3 $a \in X, \alpha \in \mathbb{R}$ such that

$$a^T m > \alpha > a^T x, \text{ for all } x \in H.$$

In particular, $a^T m > \alpha > a^T g(q)$ for all $q \in Q$. Because μ is a measure induced by a probability distribution on Q, it follows that

$$a^T m = a^T m \int_Q 1 d\mu(q) = \int_Q a^T m d\mu(q) > \int_Q a^T g(q) d\mu(q) = a^T \int_Q g(q) d\mu(q) = a^T m$$

because of property (4.5). Consequently, $a^T m > a^T m$, which is a contradiction. This proves the first statement. The second one readily follows from Lemma 4.1.

□

4.4 Size of support for optimal designs

For a subset E of $\mathbb{R}^n$, the dimension of E, denoted by $\dim(E)$, is defined to be the smallest nonnegative integer d such that up to an affine transformation $E \subseteq \mathbb{R}^d$.

Suppose that ξ is an optimal design and $M(\xi)$ is the corresponding information matrix. Thanks to Theorem 4.1, it is possible to choose ξ to be a discrete design with

at most $\frac{k(k+1)}{2}+1$ support points $q \in Q$. Let ℓ be the smallest number of points in Q needed to construct M.

$$M(\xi) = \sum_{i=1}^{\ell} \xi_i z(q_i) z(q_i)^T, \quad q_i \in Q.$$

Using arguments on the rank of a matrix, it is simple to show that $\ell \geq \left[\frac{k}{n_y}\right]$, where [ ] denotes the integer part of a real number. This yields the lower bound for ℓ.

To obtain the upper bound, we need the following lemma which is often referred to as Caratheodory's theorem of Convex Analysis.

Lemma 4.4. (Rockafellar and Wets [74, p. 55]) *If x lies in the convex hull of a set $E \subseteq \mathbb{R}^n$, then x lies in the convex hull of some subset of E that contains at most $n+1$ points. Furthermore, if $x \in \text{conv}(E)$ and $x \in \partial\,\text{conv}(E)$ - the boundary of $\text{conv}(E)$, x can be represented as a convex combination of at most n points in E.*

Proof. The first part is exactly Lemma 4.1. We prove the second statement. Suppose that $x \in \text{conv}(E) \cap \partial\,\text{conv}(E)$. By the first part, there are $x_1, x_2, ..., x_{n+1} \in E$ such that

$$x = \sum_{i=1}^{n+1} \lambda_i x_i, \ \lambda_i \geq 0, \ \sum_{i=1}^{n+1} \lambda_i = 1.$$

Since $x \in \partial\,\text{conv}(E)$, there must be some $\lambda_i = 0$ (otherwise, if all $\lambda_i > 0$, x would be in the interior of $\text{conv}(E)$). As the result, we need no more than n points of E to represent x. The proof is complete. □

Lemma 4.5. *Suppose $M(\mu)$ is an arbitrary information matrix which is nonsingular. Then there exists an information matrix $M(\xi)$ such that the number of support points of ξ is less than or equal to $k(k+1)/2$ and for the corresponding covariance matrices, it holds that*

$$\mathscr{K}(C(\xi)) \leq \mathscr{K}(C(\mu)).$$

Proof. Recall that $H = \text{conv}(g(Q))$. Set $h = \dim(H)$. If $h < \frac{k(k+1)}{2}$, then by Lemma 4.4, $M(\mu)$ can be represented as a convex combination of at most $(h+1)$ points of $g(Q)$. We can choose $M(\xi)$ as $M(\mu)$ itself. Now consider the case $h = k(k+1)/2$. Since H is convex, its interior in $\mathbb{R}^{k(k+1)/2}$ is nonempty. Define

$$\delta = \max\{\tau \geq 0, \quad \tau M(\mu) \in H\}.$$

Since $M(\mu) \in H$, $\delta \geq 1$. Also by the fact that H is compact, $\delta < \infty$. There exists a design ξ such that $M(\xi) = \delta M(\mu)$. We easily deduce that $C(\xi) = \frac{C(\mu)}{\delta}$ and

$$\mathscr{K}(C(\xi)) = \frac{\mathscr{K}(C(\mu))}{\delta} \leq \mathscr{K}(C(\mu)).$$

The definition of δ ensures that $M(\xi)$ lies on the boundary of H. By Lemma 4.4, $M(\xi)$ can be constructed by at most $k(k+1)/2$ points of $g(Q)$. This yields the desired conclusion. □

In summary, we have established the bounds of support sizes for optimal designs.

Theorem 4.2. *For any optimal design ξ, there exist ℓ points $q_1, q_2, ..., q_\ell$ in Q and positive real numbers $\xi_1, \xi_2, ..., \xi_\ell$ summing up to 1 such that*

$$M(\xi) = \sum_{i=1}^{\ell} \xi_i z(q_i) z(q_i)^T, \quad \ell \leq \frac{k(k+1)}{2}.$$

Furthermore, if the information matrix $M(\xi)$ is nonsingular, $\ell \geq \left\lfloor \frac{k}{n_y} \right\rfloor$.

4.5 Applications to OED for IVPs with inputs

In carrying out OED for the IVP (4.1), not only the time points at which measurements are performed but also inputs are chosen in order to gain as much as possible information. Thus OED for such systems can be considered as an optimal control problem (OCP), see also Sager [77]. Now we define the information matrix as a function of controls $u(\cdot)$ and probability distribution ξ on $[t_0, t_f]$, i.e.,

$$M(\xi, u(\cdot)) = \int_{t_0}^{t_f} \left(\frac{\partial \hat{\eta}}{\partial p}(x(t; t_0, x_0, u(\cdot), p), p) \right)^T \left(\frac{\partial \hat{\eta}}{\partial p}(x(t; t_0, x_0, u(\cdot), p), p) \right) d\xi(t).$$

Theorems 4.1 and 4.2 are applicable. For any ξ and $u(\cdot)$, there exists a finite support design μ on $[t_0, t_f]$ such that $M(\mu, u(\cdot)) = M(\xi, u(\cdot))$. Moreover, if ξ^*, $u^*(\cdot)$ and the corresponding trajectories $x^*(\cdot)$ as well as the information matrix M^* solve the OED problem, then there are ℓ time points $t_1, t_2, ..., t_\ell \in [t_0, t_f]$, $\ell \leq \frac{k(k+1)}{2}$ and ℓ positive numbers $\xi_1, ..., \xi_\ell$ summing up to 1 such that

$$M^* = \sum_{i=1}^{\ell} \xi_i \left(\frac{\partial \hat{\eta}}{\partial p}(x^*(t_i; t_0, x_0, u^*(\cdot), p), p) \right)^T \left(\frac{\partial \hat{\eta}}{\partial p}(x^*(t_i; t_0, x_0, u^*(\cdot), p), p) \right).$$

4.6 Constructing optimal designs with finite support

The following scheme based on Theorem 4.2 can be used to compute the optimal design with finite support:

- Set up the optimization problem: The variables comprise $n = \frac{k(k+1)}{2}$ support points $q_1, q_2, ..., q_n \in Q$ and corresponding weights $\xi_1, \xi_2, ..., \xi_n \in [0, 1]$. The covariance matrix depends on these variables, $C = C(q_1, q_2, ..., q_n, \xi_1, \xi_2, ..., \xi_n)$. With a chosen criterion, the optimization problem reads as

$$\min_{q_1, ..., q_n, \xi_1, ..., \xi_n} \mathscr{K}(C) \tag{4.6}$$

  subject to $q_i \in Q$, $\xi_i \in [0, 1]$, $i = 1, 2, ..., n$, $\sum_{i=1}^{n} \xi_i = 1$ and possibly further constraints on the experimental conditions.

- Solve the constrained optimization problem: *Sequential quadratic programming* (SQP) based methods have proven to be efficient for this kind of problems, see e.g., Körkel [46], Gill et al. [32].

Thus we are able to locate optimal support points instead of relying on preselection. Note that problem (4.6) is not convex in general and can have local minima.

4.7 Examples

Example 4.1. *Take* $Q = [0,1]$, $k = 2$ *and* $z(q) = \begin{cases} (2q \quad 1)^T & \text{if } 0 \le q < 1/2, \\ (1 \quad 2-2q)^T & \text{if } 1/2 \le q \le 1. \end{cases}$

We then have $g(q) = \begin{bmatrix} 4q^2 & 2q \\ 2q & 1 \end{bmatrix}$ *if* $0 \le q < 1/2$ *and* $g(q) = \begin{bmatrix} 1 & 2-2q \\ 2-2q & 4(1-q)^2 \end{bmatrix}$ *if* $1/2 \le q \le 1$ *where* $g(q) = z(q)z(q)^T$. *Since* $k = 2$, *in view of Theorem 4.2, the information matrix* $M(\xi)$ *corresponding to an optimal design can be constructed from* 3 *points* $q_1, q_2, q_3 \in Q$. *It is now easy to show by direct calculations that the optimal support points are contained in* $\{0; 1/2; 1\}$. *So* $M(\xi)$ *has the form*

$$M(\xi) = \xi_1 g(0) + \xi_2 g(1/2) + \xi_3 g(1), \quad 0 \le \xi_i \le 1, \ \sum_{i=1}^{3} \xi_i = 1.$$

Consider the A-criterion. We have

$$M(\xi) = \begin{bmatrix} \xi_2 + \xi_3 & \xi_2 \\ \xi_2 & \xi_1 + \xi_2 \end{bmatrix}; \quad C(\xi) = \frac{1}{(\xi_2+\xi_3)(\xi_1+\xi_2) - \xi_2^2} \begin{bmatrix} \xi_1 + \xi_2 & -\xi_2 \\ -\xi_2 & \xi_2 + \xi_3 \end{bmatrix}.$$

$$\text{Trace}(C(\xi)) = \frac{1+\xi_2}{(\xi_2+\xi_3)(\xi_1+\xi_2) - \xi_2^2} \ge \frac{4(1+\xi_2)}{(1+\xi_2)^2 - 4\xi_2^2},$$

since $(\xi_2+\xi_3)(\xi_1+\xi_2) \le \frac{(1+\xi_2)^2}{4}$. *By simple calculations we find that* Trace$(C(\xi))$ *attains its minimum* if and only if $\xi_2 = \frac{2\sqrt{3}-3}{3}$; $\xi_1 = \xi_3 = \frac{3-\sqrt{3}}{3}$. *The* unique *optimal design needs exactly* 3 *support points. Compare also the result in* [69, pp. 191-193].

Example 4.2. *We consider a Lotka-Volterra model given by*

$$\begin{cases} \dot{x}_1(t) = p_1 x_1(t) - p_2 x_1(t) x_2(t), \\ \dot{x}_1(t) = -p_3 x_1(t) + p_4 x_1(t) x_2(t), \quad t \in [0, 20], \\ x_1(0) = 100, x_2(0) = 100, \end{cases} \tag{4.7}$$

where $x_1(t), x_2(t)$ *represent the prey and predator populations at time* t. *The parameters are* $p = (p_1, p_2, p_3, p_4) \in \mathbb{R}^4$. *We use* $\eta_1(x) = x_1$; $\eta_2(x) = x_2$ *as measurement functions. In the computation below, we choose the following setting for measurements: For even time point index* i, η_1 *is applied and for odd one* i, η_2 *is applied.*

Consider OED for estimating the parameters at $p = (0.5; 0.01; 0.5; 0.02)$. *Due to Theorem 4.2, an optimal design needs at most* 10 *time points. So we apply the problem setup in Section 4.6 with the* $A-$*criterion and initialization*

$$q^0 = (3; 5; 7; 9; 11; 13; 15; 17; 18; 19), \quad \xi^0 = (\xi_{0i}), \ \xi_{0i} = 0.1 \text{ for } i = 1, 2, ..., 10.$$

The optimal solution after removing nearly zero weights (smaller than 10^{-16}*) is*

$$q^*_{1,6,8,9} = (4.42; 13.80; 17.58; 18.10); \ \xi^*_{1,6,8,9} = (0.28; 0.49; 0.10; 0.13).$$

The optimal value is $\mathscr{K}^*_A = 0.0361$. *Here we need only* 4 *time points. On the other hand, if we employ many fixed time points, namely at* $q^f_i = 0.5 + (i-1)0.5, i = 1, 2, ..., 39$ *and optimize* $\mathscr{K}_A$ *with respect to the weights* ξ^f_i *only, we arrive at the solution*

$$q^{f*}_{9,28,35,37} = (4.5; 14.0; 17.5; 18.5); \quad \xi^{f*}_{9,28,35,37} = (0.33; 0.48; 0.04; 0.15).$$

That means, an optimal design in this case needs only 4 *support points. However, the optimal value is* $\mathscr{K}^*_{Af} = 0.0373$, *which is clearly worse than* $\mathscr{K}^*_A$.

Chapter 5

Theory of Dual Control

In this chapter we formulate the Dual Control problem. The theory rests on the Dynamic Programming principle for the stochastic case and is presented in conjunction with the discrete Hamilton-Jacobi-Bellman equation (HJBE) as described in Chapter 2. In general the theoretical solution for Dual Control based on HJBE is too complicated to be implemented in practice due to the *curse of dimensionality* and nested conditional expectations. Therefore we develop approximation strategies, including Certainty Equivalence control, open-loop feedback control, Dual Control and their variants in the framework of NMPC. At the end of the chapter, we propose a new formulation for the problem of sequential experimental design and solve it with the aid of Dual Control.

5.1 Introduction

In all practical optimal control problems (OCPs), there are uncertainties in the process to be controlled. Typically some parameters are unknown and available only inexactly in the form of estimates with certain probability distributions. Consider, e.g., driving a car from A to B as fast as possible or with minimum fuel consumption. Most technical parameters of the car may be at hand but there are typically some parameters we do not know, for instance, the braking coefficient, the acceleration coefficient. The values of these parameters can have a significant effect on optimal solutions.

Feedback control is a well-known strategy to cope with uncertainties. In the framework of Nonlinear Model Predictive Control (NMPC), we estimate the parameters and the states through an estimation procedure, see Rawlings and Mayne [71]. To gain informative data for reliable estimates, Optimal Experimental Design (OED) should be carried out. Thus there are two objectives we must take into consideration. The first objective is to control the process in an optimal way specified by the objective function. We call this the *performance control* task. The second objective is to gain information about the process for reliable estimates. We call it the *information gain* task. The interplay between the two tasks is often not obvious. Sometimes they are conflicting, sometimes they support each other. Dual Control stands for strategies which assess their relationship and strike a balance in a reasonable way.

The concept of Dual Control was coined by Feldbaum [26]. He introduced Dual Control in the context of signal processing. It is worth mentioning that Dual Control was among the groundbreaking ideas of the control community, listed in the 25 seminal papers of that field, see Basar [9]. Further analysis and clarification have been carried out, for example in Åström [4], Wittenmark [86], Filatov and Unbehauen [27]. The problem of

Dual Control is traditionally solved by stochastic dynamic programming. However, this approach has experienced impediments in practical applications due to complications in computing conditional probability distributions and the *curse of dimensionality*. This has motivated the study of suitable approximations. One idea is to make use of the two fundamental properties of Dual Control (see La et al. [53]): performance control and information gain. While trying to optimize the objective function, the control should take the inaccuracy of the current estimates into account and make efforts to improve the quality of future estimates. Improved estimates help to maintain feasibility and enhance the overall performance.

In the literature, the problem of Dual Control is often tackled heuristically. The objective function is penalized by some scalarization of the covariance matrix with some weights, see Wittenmark [86], Filatov and Unbehauen [27]. OED has not been thoroughly explored. Numerical methods have not been developed adequately, especially for nonlinear processes modeled by systems of differential equations. Moreover, the combination of NMPC and Dual Control has not been investigated comprehensively.

Recent studies explored the use of OED. For example, Heirung et al. [37] proposed an approximation method for linear input-output models, Lucia and Paulen [58] presented an approach to Dual Control for robust NMPC based on scenario trees, in which parameters are supposed to assume a finite number of values. Our goal is to offer a more rigorous and viable treatment which will be introduced in Chapter 6. We approximate the *variance of the optimal nominal objective value* by a quadratic term of the covariance matrix of the corresponding OED problem and the *derivatives of the optimal nominal objective value with respect to the unknown parameters and the initial states*. This quantity can be interpreted as a *predictive variance*. Using a weighted sum of the original objective function and the predictive variance as a modified objective function, we can improve the quality of future estimates. Moreover, we provide the statistical background for our approach and probabilistic bounds for the controller performance with respect to the original objective, which are missing in the literature. The potential of our approach is illustrated by numerical examples. We remark that the computational effort for dual NMPC is demanding. The weighted covariance matrix which involves the derivatives of the states with respect to the parameters and initial states must be computed at each NMPC step. We will present in Chapter 6 efficient real-time numerical methods. They are decisive for the success of dual NMPC in practice.

The chapter is organized as follows. In Section 5.2, we give an introduction to the theory of Dual Control, including the formulation, a solution based on stochastic dynamic programming and challenges in computation. Section 5.3 presents some approximation methods to tackle those challenges. Section 5.4 introduces the problem of Dual Control for NMPC. We present new approaches to attack this problem in Section 5.5. We conclude this chapter with a formulation of sequential OED and solve it with the help of Dual Control.

5.2 Formulation of the Dual Control problem

In this section, we formulate the problem of Dual Control, present its solution via stochastic dynamic programming and analyze its behavior and applicability. We use the following conventions: Wherever upper-case letters denote random vectors (RVs), corresponding lower-case ones will denote their realizations. The reader can review Section 1.7 for the definition of conditional expectations.

We consider a discrete-time system with measurements of the form

$$\begin{cases} x_{k+1} = f_k(x_k, u_k), & k = 0, 1, 2, ..., \\ y_k = \eta_k(x_k) + \varepsilon_k, & k = 1, 2, ..., \end{cases} \tag{5.1}$$

where $x_k \in \mathbb{R}^{n_x}$ are states, $u_k \in \mathbb{U} \subseteq \mathbb{R}^{n_u}$ are controls. Often states x_k come from sampling a continuous-time system at times $t_0, t_1, ...$ The functions $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ are state transition functions. The functions $\eta_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ are measurement functions. Furthermore $y_k \in \mathbb{R}^{n_y}$ are noisy measurements and ε_k are random measurement noise. The initial value X_0 is a random vector with a known probability distribution. Note that some components of the states x_k can stay constant over k, which represent constant parameters. Also the set $\mathbb{U}$, often assumed to be closed, is used to impose control constraints. In this work we pursue a soft-constraint strategy to treat problems with state constraints, i.e., using penalty functions.

In practice, it is natural to use information obtained during operating the process to determine control actions. To this end we introduce notations that express information up to and including time $t \geq 1$,

$$\mathcal{Y}^t = (X_0, u_0, Y_1, u_1, Y_2, ..., u_{t-1}, Y_t), \quad \mathcal{Y}^0 = X_0.$$

Let $N \geq 1$. As common in stochastic control (Åström [4]), by a *control policy* we understand a sequence $U = (U_0, U_1, ..., U_{N-1})$ such that U_t is a function of $\mathcal{Y}^t$, i.e., $U_t = \mu_t(\mathcal{Y}^t)$ where $\mu_t : \mathbb{R}^{n_{yt}} \to \mathbb{U}$ with n_{yt} the dimension of $\mathcal{Y}^t$. The goal is to find in the set of all control policies a control policy U^* that solves

$$\min_U \mathbb{E} J_N(X_0, U) \tag{5.2}$$

with $J_N(x_0, u) = F_N(x_N) + \sum_{k=0}^{N-1} L_k(x_k, u_k)$ subject to (5.1) and where the expectation is taken with respect to X_0 and ε_k, $k = 0, 1, ..., N-1$. Here $L_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$ are stage cost functions and $F_N : \mathbb{R}^n \to \mathbb{R}$ is a terminal cost function.

From now on, let us denote $u_i^k = (u_i, u_{i+1}, ..., u_k)$ for integers $i \leq k$. We set

$$V(y^t, t) = \min_{u_t^{N-1}} \mathbb{E}\left[F_N(X_N) + \sum_{k=t}^{N-1} L_k(X_k, u_k) \;\middle|\; \mathcal{Y}^t = y^t \right], \quad t = 1, 2, ..., N-1.$$

The following lemma shows that $V(\mathcal{Y}^t, t)$ plays the same role as the value function in the deterministic case.

Lemma 5.1. (Åström [4, p. 270]) *Let y^t denote realizations of $\mathcal{Y}^t$. The recursive equation for $V(\mathcal{Y}^t, t)$ is given by*

$$\begin{aligned} V(y^N, N) &= \mathbb{E}\left[F_N(X_N) \mid \mathcal{Y}^N = y^N \right], \\ V(y^t, t) &= \min_{u_t} \mathbb{E}\left[L_t(X_t, u_t) + V(\mathcal{Y}^{t+1}, t+1) \mid \mathcal{Y}^t = y^t \right], \quad t = N-1, ..., 0. \end{aligned} \tag{5.3}$$

Furthermore, the optimal cost of (5.2) *is $\mathbb{E}V(\mathcal{Y}^0, 0)$.*

Remark 5.1. *The* curse of dimensionality *is well-known in dynamic programming for deterministic OCPs. If we are to apply equation* (5.3)*, the problem of dimensionality is even more dreadful. The dimension of $\mathcal{Y}^t$ rapidly grows as t increases. The computation of conditional expectations is practically impossible, especially for nonlinear systems and complicated probability distributions. Optimal solutions according to Lemma* 5.1 *are rarely known, except for few cases such as* Linear Quadratic Gaussian (LQG)*, which is presented in Section* 5.4*.*

5.3 Certainty Equivalence, open-loop feedback and Dual Control

To overcome the difficulties with dimension, we search for a way to concisely summarize all information in $\mathcal{Y}^t$ and to approximate conditional expectations. Instead of using $\mathcal{Y}^t$, we can use the best estimate of the states given $\mathcal{Y}^t$,

$$\hat{X}_t = \mathbb{E}[X_t \mid \mathcal{Y}^t].$$

This greatly reduces the dimension of the problem since the dimension of $\hat{X}_t$ is constant with respect to t. For linear systems, a Kalman filter allows us to compute this estimate recursively and efficiently. Consider the function

$$\mathcal{V}(\hat{x}_t, t) = \min_{u_t^{N-1}} \mathbb{E}\left[F_N(X_N) + \sum_{k=t}^{N-1} L_k(X_k, u_k) \;\middle|\; \hat{X}_t = \hat{x}_t \right].$$

Similar to the value function $V(\mathcal{Y}^t, t)$, it fulfills a recursive formula

$$\mathcal{V}(\hat{x}_t, t) = \min_{u_t} \mathbb{E}\left[L_t(X_t, u_t) + \mathcal{V}(\hat{X}_{t+1}, t+1) \;\middle|\; \hat{X}_t = \hat{x}_t \right], \quad t = N-1, ..., 0. \tag{5.4}$$

A fundamental question is whether $V(\mathcal{Y}^t, t) = \mathcal{V}(\hat{x}_t, t)$. This is true for the LQG in which $\hat{X}_t$ is a sufficient statistic given $\mathcal{Y}^t$, see Åström [4]. In general it does not hold.

Remark 5.2. *One can realize from* (5.4) *that the control action u_t at time t depends on the accuracy of the current estimates $\hat{X}_t$ and influences future estimates, expressed in $\mathcal{V}(\hat{X}_{t+1}, t+1)$.*

With the use of $\mathcal{V}(\hat{x}_t, t)$, the problem of dimension has been mitigated significantly. However the computation of conditional expectations is still overwhelming. Only for the linear case such as LQG can it be done exactly and practically. Otherwise, further approximations must be employed. The simplest one is to implement a *single open loop* control in which all RVs are replaced by their nominal values. More sophisticated strategies include *certainty equivalence* control, *open-loop feedback* control and *Dual Control*. We describe these schemes in the following, compare Bertsekas [10, Chapter 6].

Certainty equivalence (CE) control (nominal control) Set $t = 0$.

1. Compute $\hat{X}_t = \mathbb{E}[X_t | \mathcal{Y}^t]$.
2. Find a control sequence $\{u_t^*, u_{t+1}^*, ..., u_{N-1}^*\}$ that solves the deterministic problem

$$\min_{u_t, ..., u_{N-1}} J_N^{\text{ce}}(\hat{x}_t, u) = F_N(x_N) + \sum_{k=t}^{N-1} L_k(x_k, u_k)$$

   in which all RVs are replaced by their nominal values.
3. Apply the control $\mu_t^*(\mathcal{Y}^t) = u_t^*$ to the system.
   If $t < N-1$, take measurements y_{t+1}. Set $\mathcal{Y}^{t+1} = \{\mathcal{Y}^t, u_t^*, y_{t+1}\}$, $t = t+1$ and go to 1. If $t = N$, stop.

Nominal control does not take the inaccuracy of the estimates and the possibility of improving future estimates into account. This may badly affect the current control action. See Section 5.5 for more details. Beyond that, nominal control does not care about the quality of the estimates at next steps. The estimates after new measurements come may not be sufficiently improved. This likely further degrades the performance.

Open-loop feedback. The open-loop feedback differs from the certainty equivalence control in that the objective function in Step 2 is the expected value with respect to the RVs involved.

$$\min_{u_t,\dots,u_{N-1}} J_N^{\text{of}}(\hat{x}_t, u) = \mathbb{E}\left[F_N(X_N) + \sum_{k=t}^{N-1} L_k(X_k, u_k)\Big|\mathcal{Y}^t = y^t\right].$$

This strategy takes care of the uncertainties in $\hat{x}_t$ but the computation of conditional expectations is formidable.

Dual Control. Dual Control, also called active control, refers to strategies that take care of two goals: optimizing the objective function and probing to get information to improve the estimates in the future. There are several novel approaches to Dual Control that will be presented in depth in Chapter 6. For the completion of the exposition, we present here a Dual Control scheme that we pursue in this thesis.

Choose $\alpha \geq 0$, $0 < N_d \neq N$. Set $t = 0$, $\mathcal{Y}^0 = \{x_0\}$.

1. Compute $\hat{x}_t = \mathbb{E}[x_t|\mathcal{Y}^t]$.

2. Find a control sequence $\{u_t^*, u_{t+1}^*, ..., u_{N-1}^*\}$ that solves the deterministic problem

$$\min_{u_t,\dots,u_{N-1}} J_N^{\text{d}}(\hat{x}_t, u) = J_N(\hat{x}_t, u) + \alpha\sqrt{\delta x_t^T C^t(u, x_t)\delta x_t}. \tag{5.5}$$

where $\delta x_t = \dfrac{\partial J_N^{\text{ce}*}}{\partial \hat{x}_t}(\hat{x}_t)$ with $J_N^{\text{ce}*}(\hat{x}_t)$ is the optimal value of the $J_N^{\text{ce}}(\hat{x}_t, u)$ and $C^t(u, x_t) = (M^t)^{-1}$ with

$$M^t = \sum_{k=1}^{t+N_d} \left(\frac{\partial \eta}{\partial \theta}(x_k)\right)\left(\frac{\partial \eta}{\partial \theta}(x_k)\right)^T.$$

3. Apply the control $\mu_t^*(\mathcal{Y}^t) = u_t^*$ to the system.
If $t < N-1$, take measurements y_{t+1}. Set $\mathcal{Y}^{t+1} = \{\mathcal{Y}^t, u_t^*, y_{t+1}\}$, $t = t+1$ and go to Step 1. If $t = N-1$, stop.

It will be pointed out in Chapter 6 that the second term in (5.5) can be interpreted as the variance of the objective function caused by the uncertainty in the parameters and the initial states. Moreover, we will explore the statistical meaning of the weight α and provide guidelines to choose it appropriately.

Those schemes above represent NMPC for a fixed time horizon, i.e., batch NMPC. Similar procedures can be applied to the receding horizon control.

5.4 Linear Quadratic Gaussian

Linear Quadratic Gaussian (LQG) constitutes a fundamental result in the control theory, for which we have an explicit representation of the optimal solutions as well as a clear explanation of all terms making up the optimal objective value. The dynamic is linear with additive Gaussian noise

$$\begin{cases} x_{t+1} = A_t x_t + B_t u_t + w_t, \\ y_t = C_t x_t + v_t, \quad t = 0, 1, 2, ... \end{cases} \tag{5.6}$$

where $A_t \in \mathbb{R}^{n_x \times n_x}$, $B_t \in \mathbb{R}^{n_x \times n_u}$, $C_t \in \mathbb{R}^{n_y \times n_x}$ and

$$X_0 \sim \mathcal{N}(x_0, P_0), \quad w_t \sim \mathcal{N}(0, W_t), \quad v_t \sim \mathcal{N}(0, V_t).$$

It is also assumed that w_t, v_t are independent of X_0, and for $t > s \geq 0$, w_t and w_s, v_t and v_s are independent. The objective is to minimize the expectation of a quadratic function

$$\min_u J(u, x) = \mathbb{E}\left[X_N^T Q_N X_N + \sum_{t=1}^{N-1} \left(X_t^T Q_t X_t + u_t^T R_t u_t\right)\right],$$

subject to (5.6) where $Q_t \succeq 0, R_t \succ 0$ for $i = 1, 2, ..., N-1$ and $Q_N \succeq 0$. The reader can review Chapter 2 for the deterministic counterpart. The optimal solutions $\hat{x}_t$ and u_t can be summarized as follows (Åström [4]):

1. A Kalman filter

$$\begin{aligned} K_t &= P_t C_t^T (C_t P_t C_t^T + V_t)^{-1}, \\ P_{t+1} &= W_t + A_t P_t A_t^T - K_t (C_t P_t C_t^T + V_t) K_t^T, \\ P_0 &= \mathrm{Cov}(x_0). \end{aligned}$$

2. A Linear Quadratic Regulator

$$\begin{aligned} L_t &= (B_t^T S_{t+1} B_t + R_t)^{-1} B_t^T S_{t+1} A_t, \\ S_t &= Q_t + A_t^T S_{t+1} A_t - L_t^T (B_t^T S_{t+1} B_t + R_t) L_t, \\ S_N &= Q_N. \end{aligned}$$

3. The regulated system

$$\begin{aligned} \hat{x}_{t+1} &= A_t \hat{x}_t + B_t u_t + K_t (y_t - C_t \hat{x}_t), \\ \hat{u}_t &= -L_t \hat{x}_t, \\ \hat{x}_0 &= \mathbb{E} X_0 \end{aligned}$$

Derivation of the LGQ. We first prove the following lemmas on the expectation of the quadratic function of random variables.

Lemma 5.2. (Åström [4], p. 262) *Let Z be a random vector with $\mathbb{E}Z = z \in \mathbb{R}^{n_z}$, $\mathrm{Cov}\, Z = P$ and $Q \in \mathbb{R}^{n_z \times n_z}$. Then the following equality is valid*

$$\mathbb{E}\left(Z^T Q Z\right) = z^T Q z + \mathrm{Trace}(QP).$$

Proof. The proof is straightforward. In fact, we have

$$\begin{aligned} \mathbb{E}\left(Z^T Q Z\right) &= \mathbb{E}(Z - z)^T Q (Z - z) + \mathbb{E}\left(Z^T Q z\right) + \mathbb{E}\left(z^T Q Z\right) - z^T Q z \\ &= \mathbb{E}(Z - z)^T Q (Z - z) + z^T Q z \\ &= z^T Q z + \mathbb{E}\,\mathrm{Trace}\left(Q(Z - z)(Z - z)^T\right) \\ &= z^T Q z + \mathrm{Trace}\left(Q \mathbb{E}(Z - z)(Z - z)^T\right) \\ &= z^T Q z + \mathrm{Trace}(QP). \end{aligned}$$

□

Lemma 5.3. *In addition to the hypothesis of Lemma* 5.2*, we assume that* Y *is a random vector such that* $Z - \hat{Z}$ *and* $\hat{Z}$ *are independent, where* $\hat{Z} = \mathbb{E}\big[Z \mid Y\big]$*. Then it holds that*

$$\mathbb{E}\Big[Z^T Q Z \mid \hat{Z}\Big] = \hat{Z}^T Q \hat{Z} + \mathrm{Trace}(Q\,\mathrm{Cov}(Z - \hat{Z})).$$

Proof. Using $Z^T Q Z = (Z - \hat{Z})^T Q (Z - \hat{Z}) + 2\hat{Z}^T Q Z - \hat{Z}^T Q \hat{Z}$, we get

$$\begin{aligned}\mathbb{E}\Big[Z^T Q Z \mid \hat{Z}\Big] &= \mathbb{E}\Big[(Z - \hat{Z})^T Q (Z - \hat{Z}) \mid \hat{Z}\Big] + 2\mathbb{E}\Big[\hat{Z}^T Q Z \mid \hat{Z}\Big] - \mathbb{E}\Big[\hat{Z}^T Q \hat{Z} \mid \hat{Z}\Big] \\ &= \mathbb{E}\Big((Z - \hat{Z})^T Q (Z - \hat{Z})\Big) + 2\hat{Z} Q \mathbb{E}\big[Z \mid \hat{Z}\big] - \hat{Z}^T Q \hat{Z}, \end{aligned} \tag{5.7}$$

because $(Z - \hat{Z})^T Q (Z - \hat{Z})$ and $\hat{Z}$ are independent. On the other hand,

$$\mathbb{E}\big[Z \mid \hat{Z}\big] = \mathbb{E}\big[Z \mid \mathbb{E}[Z \mid Y]\big] = \mathbb{E}\big[Z \mid Y\big] = \hat{Z},$$

and according to Lemma 5.2

$$\mathbb{E}\Big((Z - \hat{Z})^T Q (Z - \hat{Z})\Big) = \mathrm{Trace}(Q\,\mathrm{Cov}(Z - \hat{Z}))$$

since $\mathbb{E}(Z - \hat{Z}) = 0$. Inserting these expressions into (5.7) yields the desired conclusion. □

We set

$$\mathcal{V}(\hat{x}_t, t) = \min_{u_t, u_{t+1}, \ldots, u_{N-1}} \mathbb{E}\left[X_N^T Q_N X_N + \sum_{t=1}^{N-1} X_t^T Q_t X_t + u_t^T R_t u_t \mid \hat{X}_t = \hat{x}_t\right].$$

For linear systems with Gaussian distributions, $\hat{X}_t = \mathbb{E}[X_t | \mathcal{Y}^{t-1}]$ is a sufficient statistic given $\mathcal{Y}^{t-1}$ and moreover, $X_t - \hat{X}_t$ and $\hat{X}_t$ are independent (Åström [4]). Therefore $V(y^t, t) = \mathcal{V}(\hat{x}_t, t)$. For $t = N$ and $P_N = \mathrm{Cov}(X_N - \hat{X}_N)$, we apply Lemma 5.3 to obtain

$$\mathcal{V}(\hat{x}_N, N) = \mathbb{E}\left[X_N^T Q_N X_N \mid \hat{X}_N = \hat{x}_N\right] = \hat{x}_N^T Q_N \hat{x}_N + \mathrm{Trace}(Q_N P_N). \tag{5.8}$$

We will show by induction that $\mathcal{V}(\hat{x}, t)$ has a quadratic form

$$\mathcal{V}(\hat{x}, t) = \hat{x}^T S_t \hat{x} + s_t,$$

with symmetric positive definite matrices S_t and scalars s_t. This is true for $t = N$ with

$$S_N = Q_N, \quad s_N = \mathrm{Trace}(Q_N P_N)$$

in view of (5.8). The recursive equation for $\mathcal{V}(\hat{x}_t, t)$ according to Lemma 5.1 reads as

$$\mathcal{V}(\hat{x}_t, t) = \min_{u_t} \mathbb{E}\left[X_t^T Q_t X_t + u_t^T R_t u_t + \mathcal{V}(\hat{X}_{t+1}, t+1) \mid \hat{X}_t = \hat{x}_t\right]. \tag{5.9}$$

We have due to Lemma 5.3

$$\mathbb{E}\left[X_t^T Q_t X_t \mid \hat{X}_t = \hat{x}_t\right] = \hat{x}_t^T Q_t \hat{x}_t + \mathrm{Trace}(Q_t P_t),$$

and from the well-known Kalman filter

$$\hat{X}_{t+1} = A_t \hat{X}_t + B_t u_t + K_t(Y_t - C_t \hat{X}_t),$$

$$\begin{aligned}\mathbb{E}\hat{X}_{t+1} &= A_t\hat{X}_t + B_t u_t,\\ P_{t+1} &= \operatorname{Cov}\hat{X}_{t+1} = K_t(C_tP_tC_t^T + V_t)K_t^T.\end{aligned}$$

Substituting these expressions into (5.9) we get

$$\begin{aligned}\mathcal{V}(\hat{x}_t,t) &= \min_{u_t}\mathbb{E}\Big[X_t^TQ_tX_t + u_t^TR_tu_t\\ &\qquad + (A_t\hat{X}_t + B_tu_t)^TS_{t+1}(A_t\hat{X}_t + B_tu_t) + s_{t+1} \mid \hat{X}_t = \hat{x}_t\Big]\\ &= \min_{u_t}\Big\{\hat{x}_t^TQ_t\hat{x}_t + \operatorname{Trace}(Q_tP_t) + u_t^TR_tu_t\\ &\qquad + (A_t\hat{x}_t + B_tu_t)^TS_{t+1}(A_t\hat{x}_t + B_tu_t)\\ &\qquad + \operatorname{Trace}\left(S_{t+1}K_t\left(C_tP_tC_t^T + V_t\right)K_t^T\right) + s_{t+1}\Big\}\\ &= \min_{u_t}\Big\{u_t^t(B_t^TS_{t+1}B_t + R_t)u_t + 2u_t^TB_t^TS_{t+1}A_t\hat{x}_t + \hat{x}_t^T(Q_t + A_t^TS_{t+1}A_t)\hat{x}\\ &\qquad + \operatorname{Trace}(Q_tP_t) + \operatorname{Trace}\left(S_{t+1}K_t\left(C_tP_tC_t^T + V_t\right)K_t^T\right) + s_{t+1}\Big\}\\ &= \hat{x}_t^T\left[Q_t + A_t^TS_{t+1}A_t - H^T(R_t + B_t^TS_{t+1}B_t)^{-1}H\right]\hat{x}_t\\ &\qquad + \operatorname{Trace}(Q_tP_t) + \operatorname{Trace}\left(S_{t+1}K_t(C_tP_tC_t^T + V_t)K_t^T\right) + s_{t+1}\end{aligned}$$

with $H = B_t^TS_{t+1}A_t$ as the minimization of a quadratic function, see Section 2.3. The minimum is attained by

$$\hat{u}_t = -(B_t^TS_{t+1}B_t + R_t)^{-1}B_t^TS_{t+1}A_t\hat{x}_t.$$

Setting

$$L_t = (B_t^TS_{t+1}B_t + R_t)^{-1}B_t^TS_{t+1}A_t,$$

we can write $u_t = -L_t\hat{x}_t$ and

$$\begin{aligned}\mathcal{V}(\hat{x}_t,t) =& \hat{x}_t^T\left[Q_t + A_t^TS_{t+1}A_t - L_t^T(R_t + B_t^TS_{t+1}B_t)L_t\right]\hat{x}_t\\ &+ \operatorname{Trace}(Q_tP_t) + \operatorname{Trace}\left(S_{t+1}K_t(C_tP_tC_t^T + V_t)K_t^T\right) + s_{t+1}.\end{aligned}$$

Identifying with $\mathcal{V}(\hat{x}_t,t) = \hat{x}_t^TS_t\hat{x}_t + s_t$ we obtain

$$\begin{aligned}S_t &= Q_t + A_t^TS_{t+1}A_t - L_t^T(B_t^TS_{t+1}B_t + R_t)L_t\\ &= Q_t + (A_t - B_tL_t)^TS_{t+1}(A_t - B_tL_t) + 2L_t^TB_t^TS_{t+1}A_t - 2L_t^TB_t^TS_{t+1}B_tL_t - L_t^TR_tL_t\\ &= Q_t + (A_t - B_tL_t)^TS_{t+1}(A_t - B_tL_t)\\ &\qquad + 2L_t^TB_t^TS_{t+1}A_t - 2L_t^T(B_t^TS_{t+1}B_t + R_t)L_t + L_t^TR_tL_t\\ &= Q_t + (A_t - B_tL_t)^TS_{t+1}(A_t - B_tL_t) + L_t^TR_tL_t\end{aligned}$$

and

$$s_t = \operatorname{Trace}(Q_tP_t) + \operatorname{Trace}\left(S_{t+1}K_t(C_tP_tC_t^T + V_t)K_t^T\right) + s_{t+1}.$$

Using the formula of s_t and evaluating forward the equations $\mathcal{V}(\hat{x}_t,t) = \hat{x}_t^TS_t\hat{x}_t + s_t$, one can get the optimal value in the form

$$\mathcal{V}(\hat{x}_0,0) = \hat{x}_0^TS_0\hat{x}_0 + \sum_{t=0}^{N-1}\operatorname{Trace}(Q_tP_t)$$

$$+ \sum_{t=0}^{N-1} \text{Trace}\left(S_{t+1} K_t (C_t P_t C_t^T + V_t) K_t^T\right) + \text{Trace}(Q_N P_N). \tag{5.10}$$

In order to explore the structure of the value function $\mathcal{V}(\hat{x}_0, 0)$, we will rewrite it. In the following arguments, we frequently use the fact that $\text{Trace}(AB) = \text{Trace}(BA)$ for two matrices $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times n}$. Recall that

$$\begin{aligned} P_{t+1} &= W_t + A_t P_t A_t^T - K_t (C_t P_t C_t^T + V_t) K_t^T; \\ S_t &= Q_t + A_t^T S_{t+1} A_t - L_t^T (B_t^T S_{t+1} B_t + R_t) L_t. \end{aligned}$$

Consequently,

$$\begin{aligned} P_{t+1} S_{t+1} - P_t S_t =& W_t S_{t+1} + \left(A_t P_t A_t^T S_{t+1} - P_t A_t^T S_{t+1} A_t\right) \\ &+ P_t L_t^T (B_t^T S_{t+1} B_t + R_t) L_t - P_t Q_t - K_t (C_t P_t C_t^T + V_t) K_t^T S_{t+1}. \end{aligned}$$

Taking the trace of both sides yields

$$\begin{aligned} \text{Trace}(P_{t+1} S_{t+1}) - \text{Trace}(P_t S_t) =& \text{Trace}(W_t S_{t+1}) + \text{Trace}\left(P_t L_t^T (B_t^T S_{t+1} B_t + R_t) L_t\right) \\ &- \text{Trace}(P_t Q_t) - \text{Trace}\left(K_t (C_t P_t C_t^T + V_t) K_t^T S_{t+1}\right). \end{aligned}$$

Hence

$$\begin{aligned} \text{Trace}(P_N S_N) - \text{Trace}(P_0 S_0) &= \sum_{t=0}^{N-1} \Big(\text{Trace}(P_{t+1} S_{t+1}) - \text{Trace}(P_t S_t) \Big) \\ &= \sum_{t=0}^{N-1} \text{Trace}(W_t S_{t+1}) + \sum_{t=0}^{N-1} \text{Trace}\left(P_t L_t^T (B_t^T S_{t+1} B_t + R_t) L_t\right) \\ &\quad - \sum_{t=0}^{N-1} \text{Trace}(P_t Q_t) - \sum_{t=0}^{N-1} \text{Trace}\left(K_t (C_t P_t C_t^T + V_t) K_t^T S_{t+1}\right). \end{aligned}$$

Note that $S_N = Q_N$. It follows that

$$\begin{aligned} &\text{Trace}(Q_N P_N) + \sum_{t=0}^{N-1} \text{Trace}(Q_t P_t) + \sum_{t=0}^{N-1} \text{Trace}\left(S_{t+1} K_t (C_t P_t C_t^T + V_t) K_t^T\right) \\ &= \text{Trace}(P_0 S_0) + \sum_{t=0}^{N-1} \text{Trace}(W_t S_{t+1}) + \sum_{t=0}^{N-1} \text{Trace}\left(P_t L_t^T (B_t^T S_{t+1} B_t + R_t) L_t\right). \end{aligned}$$

Comparing to (5.10), we deduce that

$$\begin{aligned} &\mathcal{V}(\hat{x}_0, 0) \\ &= \hat{x}_0^T S_0 \hat{x}_0 + \text{Trace}(P_0 S_0) + \sum_{t=0}^{N-1} \text{Trace}(W_t S_{t+1}) + \sum_{t=0}^{N-1} \text{Trace}\left(P_t L_t^T (B_t^T S_{t+1} B_t + R_t) L_t\right) \\ &= \hat{x}_0^T S_0 \hat{x}_0 + \text{Trace}(P_0 S_0) + \sum_{t=0}^{N-1} \text{Trace}(W_t S_{t+1}) + \sum_{t=0}^{N-1} \text{Trace}\left(P_t L_t^T B_t^T S_{t+1} A_t\right). \end{aligned} \tag{5.11}$$

This completes our derivation of LGQ. For more details, see Åström [4], Chapter 8.

Remark 5.3. *We now take a closer look at the structure of the optimal value* (5.11) *and analyze the contributions of various terms. The first term is the nominal cost. The second term represents the expense caused by the uncertainty in the initial values* x_0*. The third term expresses the effect of uncertainties because of disturbances* w_t*. Finally, with the presence of* P_{t+1}*, which is the covariance of the state estimates, the fourth term is the contribution of the quality of the estimates on the whole course of time. Therefore, we can reduce the overall cost by improving the quality of the estimates, e.g., reducing the covariances in a suitable sense. This inspires the study of Dual Control, which takes into account the uncertainty both at present and in future when determining control actions.*

5.5 Dual Control for NMPC

NMPC is a feedback strategy. At each NMPC step, current states and parameters are estimated by an estimation procedure, such as moving horizon estimation (MHE), extended Kalman filter (EKF), see e.g., Rawlings and Mayne [71]. These estimates and ideally their accuracy, e.g., described by the covariance matrix, are given to the next OCP. Nominal NMPC makes use of only the estimates and not their quality. This may badly affect the performance control, at least for the moment, and may even result in infeasibility. Beyond that, nominal control does not take into account the quality of the estimates in the following steps. The estimates after the arrival of new measurements may not be sufficiently improved. At the same time, by using OED, we can increase the accuracy of the future estimates. In a larger perspective, the question is to determine which information is available, and how to utilize it for our purpose, see also Bar-Shalom and Tse [8] for a relevant discussion.

The above arguments lead us to the concept of dual NMPC: A Dual Control problem which takes the accuracy of current and future estimates into consideration is formulated and solved at each NMPC step; an estimation procedure gives the estimates with significantly improved accuracy to the next OCP. Here arise interesting questions such as: What is the required accuracy of the estimates in order to have a reliable performance; and when does nominal NMPC anyway yields controls that are favorable for the estimation problem. The answers to those questions may depend on particular applications. The examples in Chapter 8 shall illustrate these issues.

As previously mentioned, the exact Dual Control solution according to Lemma 5.1 is not computable in general. Therefore, approximations must be employed, for which several approaches will be presented in Chapter 6. Recent studies on Dual Control explore the use of OED, e.g., Lucia and Paulen [58] for robust NMPC, Heirung et al. [37] for input-output systems, and La et al. [53, 54] using a sensitivity approach, demonstrating the advantages of using dual NMPC. A real-time implementation of such strategies, especially for nonlinear systems, is numerically challenging. Hence there is a need for advanced numerical methods in order to make dual NMPC applicable. With increasing computational power and advanced algorithms, e.g., Körkel [46], Bock et al. [12], this is can be efficiently tackled. In this work, we go a first step in this direction by a prototype implementation in MATLAB which allows us to test the functionality of our approaches on various problems from classics to sophisticated applications in vehicle control and chemical engineering.

5.6 Dual Control and a formulation of sequential Optimal Experimental Design

In the nonlinear case, the computed optimal design often depends on the value of the parameters. This can be seen through the following example.

Example 5.1. *Find an optimal design with respect to the A-criterion based on one measurement $\eta(x) = x$ at $t = 1$ for estimating the scalar parameter p in the following model*

$$\dot{x}(t) = -pux(t),\ t \in [0,1], \quad x(0) = x_0$$

with $x(t) \in \mathbb{R}$ is the scalar state, $u \in \mathbb{R}$ a constant control on $[0,1]$. Obviously $x(t) = x_0 e^{-put}$. The information matrix is scalar and has the form

$$M(p,u) = \left(\frac{\partial x}{\partial p}(1)\right)^2 = x_0^2 u^2 e^{-2up}.$$

For fixed p, the OED problem is to find u in order to maximize $M(p,u)$. By differentiating, one can easily obtain the optimal control is $u^* = 1/p$. Thus the optimal design depends on the value of the parameter.

The OED is solved for a particular estimate of the parameter that is different from the true value. Hence one of the fundamental questions of nonlinear OED is how to take the uncertainty in the estimates into account when designing experiments.

One approach is to model parameters as random variables with some a priori distribution. For example, suppose that p has the probability density function $\pi(p)$. The objective function for OED could be the average performance

$$J(p,u) = \mathbb{E}\mathscr{K}(C(p,u)) = \int_{\mathbb{R}^{n_p}} \mathscr{K}(C_p)\pi(p)dp. \tag{5.12}$$

The case $\pi(p) = \delta(p - p_0)$, where δ is the Dirac delta measure, corresponds to the nominal case, i.e., the parameter is assumed to be known with a deterministic value p_0. The prior distribution $\pi(p)$ often comes from expert knowledge or from previous experiments. This is plausible since estimates based on measurements with random errors are also random.

The average performance index (5.12) fails to include the possibility to reestimate the parameters when new measurements arrive. In practice, it is beneficial to do OED sequentially. Roughly speaking, we do OED for the a priori estimate p_0 of the parameters. After computing the control action u_0, applying it to the process, we take measurements at t_1. Using available measurements, we obtain a new estimate, denoted by p_1. New control actions are computed by solving an OED problem with respect to p_1. It is expected that p_1 is more accurate than p_0, hence the performance of OED will be improved. The process is repeated until no measurement is allowed. (cf. Körkel [47], Pronzato and Pázman [68])

In this manner, the problem of sequential OED, also called online OED is closely related to the Dual Control problem. However there are some difficulties. First the objective function now changes into $\mathscr{K}(C(p,u))$ which is not additive in the sense of being a sum of expressions containing u_k. Secondly we cannot obtain an LQG because the system is often nonlinear in the parameters and the objective function is not quadratic. Furthermore, so far we have the Dynamic Programming equation (5.3) only for additive objective functions. To overcome these difficulties, we shall reformulate the sequential

design problem in such a way that the Dynamic Programming principle is applicable. For the sake of a unified presentation, constant parameters p are considered as states by setting $p_{k+1} = p_k$ with $p_0 = p$. The problem of sequential OED is treated with respect to the initial value x_0.

Concretely we consider system (5.1) where we explicitly assume that $\varepsilon_k \sim \mathcal{N}(0, \mathbb{I})$. We would like to design an optimal experiment to estimate the uncertain initial value x_0. The information matrix corresponding to N measurements takes the form

$$M_N(x_0, u) = \sum_{k=1}^{N} \left(\frac{\mathrm{d}\eta_k}{\mathrm{d}x_0}(x_k) \right)^T \left(\frac{\mathrm{d}\eta_k}{\mathrm{d}x_0}(x_k) \right).$$

The goal of OED is to find the control sequence to minimize a scalar function $\mathscr{K}_1(M_N)$. We augment the system (5.1) by introducing additional variables $z_k^1, z_k^2 \in \mathbb{R}^{n_x \times n_x}$ (cf. Körkel [49])

$$\begin{aligned} z_{k+1}^1 &= \frac{\partial \eta_{k+1}}{\partial x_{k+1}}(x_{k+1}) \left(\frac{\partial x_{k+1}}{\partial x_k}(x_k) \right) z_k^1, \\ z_{k+1}^2 &= z_k^2 + (z_k^1)^T z_k^1, \quad t = 0, 1, ..., N-1 \end{aligned}$$

with the initial conditions $z_0^1 = \mathbb{I}_{n_x},\ z_0^2 = 0$. Note that z_{k+1}^1 is just $\frac{\mathrm{d}\eta_{k+1}}{\mathrm{d}x_0}$ but is expressed in the form that depends only on (x_k, z_k^1, z_k^2). The objective function can now be written in the form $J_N(u) = \mathscr{K}_1(z_N^2)$ which is a special additive function. The recursive equation for the value function of this augmented problem according to Lemma 5.3 reads as

$$\begin{aligned} V(y^N, N) &= \mathbb{E}\left\{ \mathscr{K}_1(z_N^2) \mid \mathcal{Y}^N = y^N \right\}, \\ V(y^t, t) &= \min_{u_k} \mathbb{E}\left\{ V_{k+1}(\mathcal{Y}^{t+1}) \mid \mathcal{Y}^t = y^t \right\}, \quad t = N-1, N-2, ..., 0. \end{aligned}$$

In practice, experiments may be carried out in a long or infinite time duration. We can apply the dual NMPC strategy with receding horizon. The computational expense is highly demanding due to the fact that $\mathscr{K}_1(M_N(x_0, u))$ contains sensitivities of the states with respect to the parameters, and even second derivatives for algorithms which use gradients of the objective function.

Chapter 6

Dual Control for NMPC from an OED Point of View

In Chapter 5, we pointed out the two fundamental properties of Dual Control: performance control and information gain. We shall devise in this chapter strategies that effectively explore these properties on the one hand and are computationally tractable on the other hand.

We first investigate the propagation of uncertainties along time in two scenarios: with and without measurements. This is carried out by approximating the expectation and the covariance of the states as well as of the objective value. The linear case admits an explicit representation of solutions, while approximations are needed for the nonlinear case. After that, we will present the penalization approach and the two stage-approach. Other approaches in literature are also discussed and analyzed.

The system under consideration reads as

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t), w(t)), \quad t \in [t_0, t_f], \\ x(t_0) &= x_0, \\ X(t_0) &= X_0,\ \mathbb{E}X_0 = m_0,\ \operatorname{Cov} X_0 = \Sigma_0, \\ \mathbb{E}W(t) &= 0,\ \operatorname{Cov}(W(t), W(s)) = \Sigma_w(t)\delta(t-s),\ \text{for } s \le t, \\ \operatorname{Cov}(X_0, W(t)) &= 0 \text{ for } t \ge t_0. \end{aligned} \tag{6.1}$$

Here $x(t) \in \mathbb{R}^{n_x}$ are states, $u(t) \in \mathbb{U} \subseteq \mathbb{R}^{n_u}$ are controls and $w(t) \in \mathbb{R}^{n_x}$ are random disturbances. Furthermore we write $X(t)$, $W(t)$ for random vectors together with their realizations $x(t)$ and $w(t)$, respectively. The initial values X_0 is a random vector with known expectation x_0 and known covariance Σ_0. The expectation and covariance of $W(t)$ are also known. Here $\delta(\cdot)$ denotes the Dirac delta function. Especially we assume that X_0 and $W(t)$ are uncorrelated. Note that we have not made any assumptions on the type of distributions. For the time being, we are interested in the two common statistics of random vectors: the expectation and the covariance. First we treat an OCP with the objective function of the Mayer type

$$\min \Phi(x(t_f)) \tag{6.2}$$

subject to (6.1) for a scalar function $\Phi : \mathbb{R}^{n_x} \to \mathbb{R}$.

As in Chapter 5, wherever a random vector is denoted by a capital letter, its realization will be denoted by the corresponding small one.

6.1 Propagation of uncertainties for linear systems without measurements

For simplicity as well as for motivation, we consider the linear system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + D(t)w(t). \tag{6.3}$$

where $A(t) \in \mathbb{R}^{n_x \times n_x}$, $B(t) \in \mathbb{R}^{n_x \times n_u}$ and $D(t) \in \mathbb{R}^{n_x \times n_x}$ are deterministic. Suppose that the control $u(\cdot)$ is given. Set $m(t) = \mathbb{E}X(t)$ and

$$\Sigma(t) = \mathrm{Cov}(X(t)) = \mathbb{E}[X(t) - m(t)][X(t) - m(t)]^T.$$

Lemma 6.1. (Bryson and Ho [14]) *The expectation and the covariance of $X(t)$ evolve as follows*

$$\begin{cases} \dot{m}(t) & = A(t)m(t) + B(t)u(t), \quad t \in [t_0, t_f], \\ m(t_0) & = \mathbb{E}X_0 \end{cases} \tag{6.4}$$

and

$$\begin{cases} \dot{\Sigma}(t) & = A(t)\Sigma(t) + \Sigma(t)A^T(t) + D(t)\Sigma_w D^T(t), \quad t \in [t_0, t_f], \\ \Sigma(t_0) & = \Sigma_0. \end{cases} \tag{6.5}$$

Proof. Let $F(t,s)$ denote the fundamental matrix of (6.3), i.e., $F(t,s) = \exp\int_s^t A(s)ds$, which is deterministic and satisfies

$$\frac{d}{dt}F(t,s) = A(t)F(t,s)$$

and $F(t,t) = \mathbb{I}_{n_x}$-the identity matrix of size n_x. It is well-known that $X(t)$ admits the representation

$$X(t) = F(t,t_0)X_0 + \int_{t_0}^t F(t,s)[B(s)u(s) + D(s)W(s)]ds.$$

Therefore, the mean of $X(t)$ is

$$\begin{aligned} m(t) &= F(t,t_0)x_0 + \int_{t_0}^t F(t,s)B(s)u(s)ds + \int_{t_0}^t F(t,s)D(s)\mathbb{E}W(s)ds \\ &= F(t,t_0)x_0 + \int_{t_0}^t F(t,s)B(s)u(s)ds \end{aligned}$$

since $\mathbb{E}W(s) = 0$. Taking the derivatives with respect to t yields

$$\begin{aligned} \dot{m}(t) =& \frac{d}{dt}F(t,t_0)x_0 + \int_{t_0}^t \frac{d}{dt}F(t,s)B(s)u(s)ds + F(t,t)B(t)u(t) \\ =& A(t)F(t,t_0)x_0 + A(t)\int_{t_0}^t F(t,s)B(s)u(s)ds + B(t)u(t) \\ =& A(t)\left\{F(t,t_0)x_0 + \int_{t_0}^t F(t,s)B(s)u(s)ds\right\} + B(t)u(t) \\ =& A(t)m(t) + B(t)u(t). \end{aligned}$$

This proves (6.4). We now derive the equation for the covariance. It follows from

$$X(t) - m(t) = F(t,t_0)(X_0 - x_0) + \int_{t_0}^{t} F(t,s)D(s)W(s)ds$$

that

$$\begin{aligned}
\Sigma(t) =& \mathbb{E}\left[X(t) - m(t)\right]\left[X(t) - m(t)\right]^T \\
=& F(t,t_0)\mathbb{E}\left[(X_0 - x_0)(X_0 - x_0)^T\right]F^T(t,t_0) \\
&+ \mathbb{E}\int_{t_0}^{t} F(t,s)D(s)W(s)ds \int_{t_0}^{t} W^T(s)D^T(s)F^T(t,s)ds \\
&+ \int_{t_0}^{t} F(t,t_0)\mathbb{E}\left[(X_0 - x_0)W^T(s)\right]D(s)F^T(t,s)ds \\
&+ \int_{t_0}^{t} F(t,s)D(s)\mathbb{E}\left[W(s)(X_0 - x_0)^T\right]F^T(t,t_0)ds \\
=& F(t,t_0)\Sigma_0 F^T(t,t_0) + \mathbb{E}\int_{t_0}^{t} F(t,s)D(s)W(s)ds \int_{t_0}^{t} W^T(s)D^T(s)F^T(t,s)ds.
\end{aligned}$$

We have used $\mathbb{E}\left[(X_0 - x_0)W^T(s)\right] = 0$ since X_0 and $W(s)$ are independent. Taking the derivatives of the both sides we obtain

$$\begin{aligned}
\dot{\Sigma}(t) =& \frac{d}{dt}F(t,t_0)\Sigma_0 F^T(t,t_0) + F(t,t_0)\Sigma_0\frac{d}{dt}F^T(t,t_0) \\
&+ \mathbb{E}\int_{t_0}^{t}\frac{d}{dt}F(t,s)D(s)W(s)ds \int_{t_0}^{t} W^T(s)D^T(s)F^T(t,s)ds \\
&+ \mathbb{E}\int_{t_0}^{t} F(t,s)D(s)W(s)ds \int_{t_0}^{t} W^T(s)D^T(s)\frac{d}{dt}F^T(t,s)ds \\
&+ \mathbb{E}F(t,t)D(t)W(t)\int_{t_0}^{t} W^T(s)D^T(s)F^T(t,s)ds \\
&+ \mathbb{E}\left[\int_{t_0}^{t} F(t,s)D(s)W(s)ds\right]^T W(t)D^T(t)F^T(t,t) \\
=& A(t)\left(F(t,t_0)\Sigma_0 F^T(t,t_0) + \mathbb{E}\int_{t_0}^{t} F(t,s)D(s)W(s)ds \int_{t_0}^{t} W^T(s)D^T(s)F^T(t,s)ds\right) \\
&+ \left(F(t,t_0)\Sigma_0 F^T(t,t_0) + \mathbb{E}\int_{t_0}^{t} F(t,s)D(s)W(s)ds \int_{t_0}^{t} W^T(s)D^T(s)F^T(t,s)ds\right)A^T(t) \\
&+ D(t)\int_{t_0}^{t}\mathbb{E}[W(t)W^T(s)]D^T(s)F^T(t,s)ds + \int_{t_0}^{t} F(t,s)D(s)\mathbb{E}[W(s)W^T(t)]dsD^T(t).
\end{aligned}$$

The two terms on the last line can be computed by means of integrals with the Dirac delta function. In fact, we have

$$\int_a^b g(s)\delta(b-s)ds = \frac{g(b)}{2}$$

for any function $g(s)$ and real numbers $a < b$. Therefore

$$D(t)\int_{t_0}^{t}\mathbb{E}[W(t)W^T(s)]D^T(s)F^T(t,s)ds = D(t)\int_{t_0}^{t}\Sigma_w(t)\delta(t-s)(s)D^T(s)F^T(t,s)ds$$

$$= \frac{1}{2}D(t)\Sigma_w(t)D^T(t)F^T(t,t) = \frac{1}{2}D(t)\Sigma_w(t)D^T(t)$$

Similarly

$$\int_{t_0}^{t} F(t,s)D(s)\mathbb{E}[W(s)W^T(t)]dsD^T(t) = \frac{1}{2}D(t)\Sigma_w(t)D^T(t)$$

All together, we obtain

$$\dot{\Sigma}(t) = A(t)\Sigma(t) + \Sigma(t)A^T(t) + D(t)\Sigma_w(t)D^T(t).$$

The proof is complete. □

It is worth noting that the covariance of $X(t)$ is independent of its expectation and the control $u(\cdot)$. Now our interest is to use some scalarization of this covariance matrix. To this end, we first prove a simple lemma.

Lemma 6.2. *Suppose that $Y(t_f) = a^T X(t_f)$ where $a \in \mathbb{R}^{n_x}$ is a given deterministic vector. Then the variance of $Y(t_f)$ can be computed via*

$$\operatorname{Cov} Y(t_f) = \lambda^T(t_0)\Sigma_0\lambda(t_0) + \int_{t_0}^{t_f} \lambda^T(t)D(t)\Sigma_w(t)D^T(t)\lambda(t)dt, \tag{6.6}$$

where $\lambda(t)$ satisfies

$$\dot{\lambda}(t) = -A^T(t)\lambda(t), \quad t \in (t_0, t_f), \quad \lambda(t_f) = a. \tag{6.7}$$

Proof. Let $\lambda : [t_0, t_f] \to \mathbb{R}^{n_x}$ be defined as in (6.7). We have

$$\operatorname{Cov} Y(t_f) = \operatorname{Cov}(a^T X(t_f)) = a^T \operatorname{Cov}(X(t_f))a = \lambda(t_f)^T\Sigma(t_f)\lambda(t_f).$$

In this form, using (6.5) we deduce that

$$\begin{aligned}
\operatorname{Cov} Y(t_f) =& \lambda^T(t_0)\Sigma(t_0)\lambda(t_0) + \int_{t_0}^{t_f} \frac{d}{dt}\left(\lambda^T(t)\Sigma(t)\lambda(t)\right) dt \\
=& \lambda^T(t_0)\Sigma_0\lambda(t_0) + \int_{t_0}^{t_f} \dot{\lambda}^T(t)\Sigma(t)\lambda(t)dt + \int_{t_0}^{t_f} \lambda^T(t)\Sigma(t)\dot{\lambda}(t)dt \\
&+ \int_{t_0}^{t_f} \lambda^T(t)\dot{\Sigma}(t)\lambda(t)dt \\
=& \lambda^T(t_0)\Sigma_0\lambda(t_0) + \int_{t_0}^{t_f} \dot{\lambda}^T(t)\Sigma(t)\lambda(t)dt + \int_{t_0}^{t_f} \lambda^T(t)\Sigma(t)\dot{\lambda}(t)dt \\
&+ \int_{t_0}^{t_f} \lambda^T(t)\left[A(t)\Sigma(t) + \Sigma(t)A^T(t) + D(t)\Sigma_w(t)D^T(t)\right]\lambda(t)dt \\
=& \lambda^T(t_0)\Sigma_0\lambda(t_0) + \int_{t_0}^{t_f} \lambda^T(t)D(t)\Sigma_w(t)D^T(t)\lambda(t)dt \\
&+ \int_{t_0}^{t_f} \left[\dot{\lambda}(t) + A^T(t)\lambda(t)\right]^T \Sigma(t)\lambda(t)dt + \int_{t_0}^{t_f} \lambda^T(t)\Sigma(t)\left[\dot{\lambda}(t) + A^T(t)\lambda(t)\right] dt.
\end{aligned}$$

Because $\dot{\lambda}(t) = -A^T(t)\lambda(t), \quad t \in [t_0, t_f]$, the formula for $\operatorname{Cov} Y(t_f)$ reduces to (6.6). This completes the proof. □

To approximate the variance of the objective function $\Phi(x(t_f))$ of (6.2), one can use a linear approximation of $\Phi(x(t_f))$ around $m(t_f) = \mathbb{E}X(t_f)$,

$$\Phi(X(t_f)) \approx \Phi_a(X(t_f)) = \Phi(m(t_f)) + \frac{\partial \Phi}{\partial x}(m(t_f))(X(t_f) - m(t_f)),$$

and obtain

$$\operatorname{Cov} \Phi(X(t_f)) \approx \operatorname{Cov} Y(t_f) \tag{6.8}$$

where $Y(t_f) = \frac{\partial \Phi}{\partial x}(m(t_f))X(t_f)$. Then $\lambda(t)$ according to (6.7) satisfies

$$\dot{\lambda}(t) = -A^T(t)\lambda(t),\ t \in (t_0, t_f); \quad \lambda(t_f) = \left(\frac{\partial \Phi}{\partial x}\right)^T (m(t_f))$$

which is exactly the adjoint variables that fulfill the PMP.

For objective functions containing both a Mayer term and a Lagrange term such as

$$J(x_0, u) = \Phi(x(t_f)) + \int_{t_0}^{t_f} L(t, x(t), u(t))dt,$$

one can introduce auxiliary variables to arrive at the Mayer case and then employ the above approximation. In fact, set $z(t) = (x(t), x_a(t))$ with $x_a(t)$ satisfies

$$\dot{x}_a(t) = L(t, x(t), u(t)), \quad x_a(t_0) = 0.$$

The new objective function is of Mayer type like (6.2)

$$J_z(x_0, u) = \Phi(x(t_f)) + x_a(t_f).$$

6.2 Propagation of uncertainties for linear systems with measurements

In addition to (6.3) we consider measurements of the form

$$y(t) = C(t)x(t) + v(t),$$

where $v(t) \sim \mathcal{N}(0, R(t))$, $R(t) \succ 0$. We shall use the PMP to derive the Kalman-Bucy filter. For this purpose, we consider an optimal linear unbiased estimator of the form

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + K(t)\left[y(t) - C(t)\hat{x}(t)\right].$$

with the initial condition $\hat{X}(t_0) = X_0$. Here $K(t)$ are gain matrices that need to be determined. The error estimate $e(t) = x(t) - \hat{x}(t)$ satisfies

$$\begin{aligned}\dot{e}(t) &= A(t)(x(t) - \hat{x}(t)) - K(t)[C(t)x(t) + v(t) - C(t)\hat{x}(t)] \\ &= [A(t) - K(t)C(t)](x(t) - \hat{x}(t)) - K(t)v(t) \\ &= [A(t) - K(t)C(t)]e(t) - K(t)v(t),\end{aligned}$$

and $e(t_0) = 0$. It follows that

$$\frac{d}{dt}\mathbb{E}e(t) = [A(t) - K(t)C(t)]\mathbb{E}e(t) + K(t)\mathbb{E}v(t) = [A(t) - K(t)C(t)]\mathbb{E}e(t).$$

This is a linear system with the initial condition $\mathbb{E}e(t_0) = 0$. Therefore $\mathbb{E}e(t) = 0$ for $t \geq t_0$. Similar to the proof of Lemma 6.1, we have that, the error covariance $\Sigma(t) = \mathbb{E}e(t)e^T(t)$, which is also the covariance of the state estimates, fulfills

$$\dot{\Sigma}(t) = [A(t) - K(t)C(t)]\Sigma(t) + \Sigma(t)[A(t) - K(t)C(t)]^T + K(t)R(t)K^T(t). \tag{6.9}$$

The gain matrix $K(t)$ is now chosen to minimize the expectation of the squares error

$$\min_{K(t)} \mathbb{E}e^T(t)e(t) = \operatorname{Trace}\mathbb{E}(e(t)e^T(t)) = \operatorname{Trace}\Sigma(t).$$

subject to (6.9). Considering $K(t)$ as a control function, we will use the PMP to find an optimal $K^*(t)$. The Hamiltonian of this optimal control problem is

$$H(\Sigma, K, t, \Lambda) = \sum_{i,j} \Lambda_{ij}F_{ij} = \operatorname{Trace}(F\Lambda)$$

where F is the right hand side of (6.9). The adjoint variables $\Lambda(t)$ satisfy

$$\dot{\Lambda}(t) = -\frac{\partial H}{\partial \Sigma} = -\Lambda^T(t)[A(t) - K(t)C(t)] - [A(t) - K(t)C(t)]^T\Lambda(t)$$

with the final time conditions $\Lambda(t_f) = \mathbb{I}_n$. One can easily see that $\Lambda(t)$ is positive definite. Applying the PMP, we obtain

$$\begin{aligned} 0 = \frac{\partial H}{\partial K} = &- \Lambda(t)\Sigma(t)C^T(t) - \Lambda^T(t)\Sigma(t)C^T \\ &+ \Lambda(t)K(t)R(t) + \Lambda^T(t)K(t)R(t). \end{aligned}$$

It follows from the positive symmetric property of $\Lambda(t)$ that

$$-2\Lambda(t)\Sigma(t)C^T(t) + 2\Lambda(t)K(t)R(t) = 0.$$

Therefore

$$K(t) = \Sigma(t)C^T(t)R^{-1}(t).$$

The ODE (6.9) for the error covariance can be rewritten as

$$\dot{\Sigma}(t) = A(t)\Sigma(t) + \Sigma(t)A^T(t) - \Sigma(t)C^T(t)R^{-1}(t)C(t)\Sigma(t)$$

with the initial conditions $\Sigma(0) = \operatorname{Cov} X_0$. In summary, we have proved

Lemma 6.3. (Bryson and Ho [14]) *The optimal linear unbiased estimate $\hat{x}(t)$ of $x(t)$ is determined by*

$$\begin{aligned} \dot{\hat{x}}(t) &= A(t)\hat{x}(t) + B(t)u(t) + K(t)(y(t) - C(t)\hat{x}(t)), \quad \hat{x}(0) = \mathbb{E}X_0; \\ \dot{\Sigma}(t) &= A(t)\Sigma(t) + \Sigma(t)A^T(t) - \Sigma(t)C^T(t)R^{-1}(t)C(t)\Sigma(t), \quad \Sigma(0) = \operatorname{Cov} X_0, \end{aligned}$$

with the gain matrix $K(t) = \Sigma(t)C^T(t)R^{-1}(t)$.

Remark 6.1. *For nonlinear systems* (6.1)*, we perform a linearization around some reference trajectory, say $\bar{x}(t), \bar{u}(t)$ and $w_0(t) = 0$. System* (6.1) *is approximated by the linear system* (6.3) *with*

$$A(t) = \frac{\partial f}{\partial x}(t, \bar{x}(t), \bar{u}(t), 0),\ B(t) = \frac{\partial f}{\partial u}(t, \bar{x}(t), \bar{u}(t), 0),\ D(t) = \frac{\partial f}{\partial w}(t, \bar{x}(t), \bar{u}(t), 0).$$

The variance of $\Phi(x(t_f))$ is approximated as in (6.8) *accordingly.*

In the remainder of this chapter, we present several practical approaches to Dual Control for nonlinear models. They rest on approximations of nonlinearity and covariance matrix. One of the conventional methods is to linearize the nonlinear system and apply Linear Quadratic Gaussian (LQG). Other sophisticated methods are based on OED. The reader can review Chapter 4 for an introduction to OED and especially Section 4.5 for OED with controls.

6.3 Nonlinear Systems and Linear Quadratic Gaussian

Consider a nonlinear, deterministic system

$$\dot{x} = g(t, x(t), u(t)),\ t \in (t_0, t_f) \tag{6.10}$$

with the initial condition $x(0) = x_0$. The desired control $\bar{u}(t)$ and the desired trajectory $\bar{x}(t)$ are chosen. The goal is to design a control policy so that (6.10) tracks these reference solutions. To this end, we introduce the variables $\delta x(t)$ and $\delta u(t)$ which represent the correction to the reference solutions. Set

$$A(t) = \frac{\partial g}{\partial x}(t, \bar{x}(t), \bar{u}(t)), \quad B(t) = \frac{\partial g}{\partial u}(t, \bar{x}(t), \bar{u}(t)).$$

If present, the measurement function $\eta(x_t)$ is also linearized with $C(t) = \frac{\partial \eta}{\partial x}(\bar{x}(t))$. The weighting matrices $Q(t) \succeq 0, R(t) \succ 0$ are selected. We then solve the LQG problem

$$\min_{\delta u} J(\delta u) = \Phi(\delta x(t_f)) + \int_{t_0}^{t_f} \left[(\delta x(t))^T Q(t) \delta x(t) + (\delta u(t))^T R(t) \delta u(t)\right] dt$$

subject to

$$\delta \dot{x}(t) = A(t)\delta x(t) + B(t)\delta u(t) + w(t),$$

together with the measurements

$$\delta y(t) = C(t)\delta x(t) + v(t).$$

Consequently, we obtain an optimal control law $\delta u^*(t) = -L(t)\delta x^*(t)$. The actual controls that are applied to the real system are $u(t) = \bar{u}(t) + \delta u^*(t)$.

In order to implement this scheme in practice, we need effective methods to compute the derivatives of $f(\cdot)$ and $\eta(\cdot)$ with respect to states and control. To handle this problem one can use automatic differentiation techniques (Walther and Griewank [85]). Consider the case when $\bar{x}(t), \bar{u}(t)$ are constant, i.e., $\bar{x}(t) \equiv \bar{x}_0 \in \mathbb{R}^{n_x}$, $\bar{u}(t) \equiv \bar{u}_0 \in \mathbb{R}^{n_u}$. We then obtain a time-invariant system for which these derivatives are evaluated only at $(\bar{x}_0, \bar{u}_0)$. The computational effort is thus drastically reduced.

Random vectors $w(t), v(t)$ might be chosen based on linearization errors. The choice of the weighting matrices may depend on particular situations. For more details on the practical use of LQG, consult Athans [5].

6.4 New approaches to Dual Control

In this section, we propose novel approaches to attack the Dual Control problem. Based on the sensitivities of the optimal objective value with respect to parameters and initial states, we approximate its variance and try to reduce it. In connection with the

soft-constraint strategy, reduction of the variance can increase the probability of state constraint fulfillment.

For numerical treatments of OCPs, we pursue a direct approach. This means, we first discretize controls and states, transform the problem into nonlinear programs, then use state-of-the-art numerical methods to solve the discretized problem. Suppose that the OCP in consideration reads as

$$J(x_0, u) = \Phi(x(t_f)) + \int_{t_0}^{t_f} L(t, x(t), u(t))dt, \tag{6.11}$$

subject to

$$\dot{x}(t) = f(t, x(t), u(t)), \quad t \geq 0; \quad x(t_0) = x_0. \tag{6.12}$$

We can also have a measurement function $\eta : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$ together with noisy measurements $y(t) \in \mathbb{R}^{n_y}$ given by

$$y(t) = \eta(x(t)) + \varepsilon(t)$$

where $\varepsilon(t)$ is measurement noise, assumed to be white noise.

Let $\tau \geq 0$ and $u(\cdot)$ be given. Denote by $x(t; \tau, x^0, u(\cdot))$ the solution of (6.12) with $x(\tau) = x^0$. Consider a sampling grid $t_0 < t_1 < t_2 < ... < t_N = t_f$. As explained in Section 1.1, the continuous time OCP (6.11)–(6.12) can be transformed into a discrete time OCP of the form

$$\min_{u_0,...,u_{N-1}} F_N(x_N) + \sum_{k=0}^{N-1} L_k(x_k, u_k) \tag{6.13}$$

subject to

$$x_{k+1} = f_k(x_k, u_k), \quad k = 0, 1, 2, ..., N-1 \tag{6.14}$$

with x_0 given. In addition, measurements at sampling points have the form

$$y_k = \eta(x_k) + \varepsilon_k, \quad k = 1, 2, ..., N-1$$

where $\varepsilon_k = \varepsilon(t_k)$. We assume without loss of generality that $\varepsilon_k \sim \mathcal{N}(0, \mathbb{I})$.

For this reason, and also in agreement with Chapter 5, Dual Control methods will be presented in discrete time form. Recall that u_i^ℓ with integers i, ℓ denotes $(u_i, ..., u_\ell)$.

For readability we recall basic facts about OED. Let $i \geq 0$ be fixed. Let $i \geq 0$ be fixed. Suppose that u_i^{i+N-1} is given and the states x_i are to be estimated. Furthermore, there is prior information about x_i in form of an initial estimate x_i^0 together with a prior covariance matrix Σ_i, assumed to be positive definite. The least-squares (LS) method for parameter estimation based on measurements $y_k,\ k = i+1, i+2, ..., i+N$ reads as

$$\min_{x_i} \frac{1}{2}(x_i - x_i^0)^T \Sigma_i^{-1}(x_i - x_i^0) + \frac{1}{2} \sum_{k=i+1}^{i+N} \|y_k - \eta(x_k)\|^2 . \tag{6.15}$$

subject to (6.14). The information matrix is defined to be

$$M(x_i, u_i^{i+N-1}) = \Sigma_i^{-1} + \sum_{k=i+1}^{i+N} \left(\frac{\mathrm{d}\eta}{\mathrm{d}x_i}(x_k)\right)^T \left(\frac{\mathrm{d}\eta}{\mathrm{d}x_i}(x_k)\right). \tag{6.16}$$

Since Σ_i is positive definite, so is $M(x_i, u_i^{i+N-1})$. We define the covariance matrix by

$$\Sigma(x_i, u_i^{i+N-1}) = M^{-1}(x_i, u_i^{i+N-1}).$$

OED aims to minimize some scalar function $\mathscr{K}(\Sigma(x_i, u_i^{i+N-1}))$ of the covariance matrix by choosing the control u suitably, see La et al. [55],

$$\min_{u_i^{i+N-1}} \mathscr{K}(\Sigma(x_i, u_i^{i+N-1})). \tag{6.17}$$

Popular criteria include the *A-criterion*: $\mathscr{K}(\Sigma) = \frac{1}{n_x}\text{Trace}(\Sigma)$; the *D-criterion*: $\mathscr{K}(\Sigma) = (\det(\Sigma))^{\frac{1}{n_x}}$. In the following, we pay special attention to the *G-criterion* in square root form: $\mathscr{K}(\Sigma) = \sqrt{g^T \Sigma g}$ with a given vector $g \in \mathbb{R}^n$.

Let $0 \leq i < \ell$ be fixed. Associated with (6.14), the following OCP is considered

$$\min_u J(x_i, u_i^\ell, i, \ell) = F_\ell(x_\ell) + \sum_{k=i}^{\ell} L_k(x_k, u_k). \tag{$P(x_i, i, \ell)$}$$

Assume that for each x_i, OCP $(P(x_i, i, \ell))$ is solvable with the optimal value $J^*(x_i, i, \ell)$. For brevity, we write $J(x_i, u_i^\ell) = J(x_i, u_i^\ell, i, \ell)$ and $J^*(x_i) = J^*(x_i, i, \ell)$ when i, ℓ are already specified. Regarding the uncertainty in x_i, we would like to strike a balance between the objective of OCP $(P(x_i, i, \ell))$ and the objective of OED (6.17). To this end, we compute the sensitivities of the optimal value with respect to the initial states

$$\delta x_i = \nabla_{x_i} J^*(x_i).$$

We consider X_i as a random vector with mean x_i and covariance Σ_i. Using a Taylor expansion of $J^*(X_i)$ around x_i, we obtain,

$$J^*(X_i) - J^*(x_i) = \delta x_i^T (X_i - x_i) + R(\|X_i - x_i\|),$$

where the remainder $R(r) = o(r)$. It follows that

$$\begin{aligned}
\text{Var} J^*(X_i) =& \mathbb{E}\big[J^*(X_i) - J^*(x_i)\big]^2 = \mathbb{E}\big[\delta x_i^T (X_i - x_i)(X_i - x_i)^T \delta x_i\big] \\
&+ 2\mathbb{E}\Big(\delta x_i^T (X_i - x_i) R(\|X_i - x_i\|)\Big) + \mathbb{E} R(\|X_i - x_i\|)^2 \\
=& \delta x_i^T \mathbb{E}\big[(X_i - x_i)(X_i - x_i)^T\big] \delta x_i + o(\mathbb{E}\,\|X_i - x_i\|^2) \\
=& \delta x_i^T \Sigma_i \delta x_i + o(\mathbb{E}\,\|X_i - x_i\|^2) \approx \delta x_i^T \Sigma_i \delta x_i.
\end{aligned}$$

The third equality holds because of the Cauchy-Schwarz inequality for expectations and the Lebesgue dominated convergence for passing the limit through the integral sign. Note that Σ_i is the initial covariance of X_i. In the framework of NMPC, we apply the first N_c elements of the computed control sequence. After measurements and estimation, a new covariance matrix $\Sigma(x_i, u_i^{i+N_c-1})$ for x_i can be obtained. We are interested in the variance of the optimal objective value with respect to the new estimates, which can be interpreted as a *predictive variance*. The *predictive standard deviation* is then approximated by

$$\sqrt{\text{Var}\, J^*(X_i)} \approx \sqrt{\delta x_i^T \Sigma(x_i, u_i^{i+N_c-1}) \delta x_i}.$$

Our idea is to reduce the predictive standard deviation, which in turn reduces the covariance of the estimated parameters. This is closely related to the G-criterion with $g = \delta x_i$. In addition, if $\delta x_i^T \Sigma(x_i, u_i^{i+N_c-1}) \delta x_i$ is sufficiently small, i.e., the objective value is insensitive to the uncertainty of the initial states, we can safely ignore the predictive standard deviation.

6.4.1 Penalization approach

We can now balance the two objectives of Dual Control: The first objective is to minimize the nominal cost function. The second objective is to reduce the predictive standard deviation caused by the uncertain initial states. Guided by *bi-objective optimization*, we consider the following weighted sum as a modified objective function

$$J_d(x_i, u_i^\ell) = J(x_i, u_i^\ell) + \alpha\sqrt{\delta x_i^T \Sigma(x_i, u_i^\ell)\delta x_i}$$

where α is a nonnegative constant. Using the analysis above, we can give a statistical background for this approach and bounds for the actual controller performance with respect to the original objective together with a guideline to choose α appropriately based on the distribution of $J(X_i, u_i^\ell)$. For instance, assume that $J(X_i, u_i^\ell)$ is a Gaussian random variable. As $\sqrt{\delta x_i^T \Sigma(x_i, u_i^\ell)\delta x_i}$ is an approximation of the standard deviation of $J(X_i, u_i^\ell)$, the weight α corresponds to a *quantile* of its distribution. To each $0 < \gamma < 1$, we choose α so that $J_d(x_i, u_i^\ell)$ accounts for a *confidence interval* with confidence level γ. Statistically, this means that if we are to draw many random values of x_i from the distribution of X_i, then

$$J(\bar{x}_i, u_i^\ell) \leq J_d(x_i, u_i^\ell) \quad \text{with probability } \gamma \tag{6.18}$$

for any feasible u_i^ℓ, where $\bar{x}_i$ denotes the true value of x_i (consult e.g., Rice [73] for the meaning of a confidence interval). Suppose that the OCP with the objective function $J_d(x_i, u_i^\ell)$ has a solution $u_i^{\ell d}$ with the optimal value $J_d^*(x_i)$. Similarly OCP $(P(\bar{x}_i, i, \ell)$ has a solution $u_i^{\ell *}$ with the optimal value $J^*(\bar{x}_i)$. Choosing $u_i^\ell = u_i^{\ell d}$ in (6.18) yields

$$J(\bar{x}_i, u_i^{\ell d}) \leq J_d(x_i, u_i^{\ell d}) = J_d^*(x_i) \quad \text{with probability } \gamma.$$

Further it follows from the optimality of $u_i^{\ell *}$ that

$$J^*(\bar{x}_i) = J(\bar{x}_i, u_i^{\ell *}) \leq J(\bar{x}_i, u_i^{\ell d}) \leq J_d^*(x_i) \quad \text{with probability } \gamma.$$

In other words, the interval $[J^*(\bar{x}_i), J_d^*(x_i)]$ is a confidence interval with confidence level γ for the realized controller performance $J(\bar{x}_i, u_i^{\ell d})$. In a special case, $\gamma = 0.95$ leads to the choice $\alpha = 1.96$. Thus, $J_d^*(x_i)$ provides a reliable upper bound for $J(\bar{x}_i, u_i^{\ell d})$. Moreover increasing the values of α will increase the reliability but may worsen this upper bound. We remark that the distribution of $J(X_i, u_i^\ell)$ might be more complicated than Gaussian, hence a suitable choice of α needs to be investigated carefully in the general case.

Penalization dual NMPC algorithm Choose integers $0 < N_c \leq N_d \leq N$, tolerance `tol` > 0, weight $\alpha > 0$. The estimate x_0 is given together with an invertible prior covariance matrix Σ_0. Set $i = 0$.

1. Solve OCP $P(x_i, i, i+N)$ to get the nominal optimal value $J^*(x_i)$ and nominal optimal control sequence $u^* = (u_i^*, ..., u_{i+N-1}^*)$. Compute the sensitivity

$$\delta x_i = \nabla_{x_i} J^*(x_i).$$

2. (Dual OCP) If $\delta x_i^T \Sigma_i \delta x_i \leq \texttt{tol}$, set $\bar{u} = (u_i^*, ..., u_{i+N-1}^*)$. Go to 3.
If $\delta x_i^T \Sigma_i \delta x_i > \texttt{tol}$, find a control sequence $\bar{u} = (u_i^*, ..., u_{i+N-1}^*)$ that solves the following OCP

$$\min_{u_i^{i+N-1}} J(x_i, u_i^{i+N-1}) + \alpha \sqrt{\delta x_i^T \Sigma(x_i, u_i^{i+N_d-1}) \delta x_i}, \tag{6.19}$$

subject to (6.14) starting from $k = i$. Here $\Sigma(x_i, u_i^{i+N_d-1})$ is computed based on only N_d new measurements, $\Sigma(x_i, u_i^{i+N_d-1}) = M^{-1}(x_i, u_i^{i+N_d-1})$ with

$$M(x_i, u_i^{i+N_d-1}) = \Sigma_i^{-1} + \sum_{k=i+1}^{i+N_d} \left(\frac{\mathrm{d}\eta}{\mathrm{d}x_i}(x_k) \right) \left(\frac{\mathrm{d}\eta}{\mathrm{d}x_i}(x_k) \right)^T. \tag{6.20}$$

3. Apply $\bar{u}_i^{i+N_c-1} = (u_i^*, ..., u_{i+N_c-1}^*)$ to the system.

4. (MHE) Take new measurements $y_{i+1}, y_{i+2}, ..., y_{i+N_c}$ and estimate the states $\hat{x}_i$ by solving a least-squares problem of the form

$$\min_{\hat{x}_i} \frac{1}{2} (\hat{x}_i - x_i)^T \Sigma_i^{-1} (\hat{x}_i - x_i) + \frac{1}{2} \sum_{k=i+1}^{i+N_c} \|y_k - \eta(\hat{x}_k)\|^2 \tag{6.21}$$

subject to (6.14), i.e., $\hat{x}_{k+1} = f_k(\hat{x}_k, u_k^*)$ for $k = i, ..., i + N_c - 1$. Approximate the covariance of $\hat{x}_{i+N_c}$ by

$$\hat{\Sigma}_{i+N_c} = \left(\frac{\mathrm{d}x_{i+N_c}}{\mathrm{d}x_i} \right)^T \Sigma(\hat{x}_i, \bar{u}_i^{i+N_c-1}) \frac{\mathrm{d}x_{i+N_c}}{\mathrm{d}x_i}.$$

Set $i = i + N_c$, $x_i = \hat{x}_i$, $\Sigma_i = \hat{\Sigma}_i$.
Stop, if the final time is met, else go to 1.

Henceforth, we call N the *prediction horizon*, N_c the *control horizon* and N_d the *dual horizon*. The reason for choosing $N_c \leq N_d \leq N$ can be clarified as follows. At each NMPC step, N_c measurements are to be obtained. The number of measurements N_d used for OED, which are supposed to be taken, should be greater than or equal to the actual taken measurements N_c but not greater than the prediction horizon N.

Remark 6.2. *There are variants for MHE depending on the allowed sample sizes and on the way to weight the a priori covariance matrix Σ_i, see e.g., Kühl et al.* [50].

6.4.2 Two-stage approach

This approach is different from the penalization approach in that, instead of weight α, we choose an *objective investment* $0 < \beta < 1$, usually small and instead of OCP (6.19) in Step 2, we solve the following OCP

$$\min_{u_i^{i+N-1}} \delta x_i^T \Sigma(x_i, u_i^{i+N-1}) \delta x_i, \tag{6.22}$$

subject to (6.14) and

$$J(u_i^{i+N-1}, x_i) \leq (1 + \beta \mathrm{sign}(J^*(x_i))) J^*(x_i) \tag{6.23}$$

or

$$J(u_i^{i+N-1}, x_i) \leq J^*(x_i) + \beta. \tag{6.24}$$

When $|J^*(x_i)|$ is much smaller than 1, the choice of constraint (6.24) may be more appropriate. Otherwise, (6.23) is preferable. Moreover, it is expected that constraint (6.23) and (6.24) are active in the solution of the corresponding OCP.

Remark 6.3. *The above algorithms can be naturally extended for batch NMPC.*

6.5 Other approaches

A naive path to Dual Control is to deal with OED and optimal control separately. This means that for the first several steps, we minimize a positive scalar function of the covariance matrix, i.e.,

$$\min_{u_i^{i+N-1}} \mathscr{K}(\Sigma(x_i, u_i^{i+N-1})).$$

We call this approach *naive* Dual Control because it spends all the effort at the beginning to handle the uncertainties in the unknown parameters and states. It might be expected to be robust in the sense that the estimates get accurate quickly, ensuring constraints with high probability. However, this approach completely ignores the performance control, and hence, is likely to worsen the overall performance. A slight modification is to solve

$$\min_{u_i^{i+N-1}} J(x_i, u_i^{i+N-1}) + \alpha \mathscr{K}(\Sigma(x_i, u_i^{i+N-1}))$$

with some positive constant α. A big α expresses our willingness to sacrifice the original objective in exchange for more information. The choice of α is critically important.

An obvious shortcoming of approaches which use the conventional criteria from OED is that they pay no attention to the effect of unknown parameters on the objective function. They aim at achieving estimates with excessively high accuracy that might not be necessary for the time being or even when some of the parameters have no effect on the objective function for a long period. The effort to improve the accuracy of estimates is used at the wrong time and the information gain task is misused. This may degrade the overall performance.

The above mentioned drawback can be avoided by making use of sensitivity analysis. This helps to determine when and how strongly uncertain parameters affect the objective function. For this purpose, we compute the derivatives of the optimal objective value with respect to the parameters. This is the underpinning for our approaches presented in Section 6.4.

Chapter 7

Partial Stability for Nonlinear Model Predictive Control

While the theory of stability for nonlinear model predictive control (NMPC) has been under extensive study, partial stability is apparently overlooked. In this chapter we apply the concept of partial stability in order to extend several results for the stability analysis of NMPC and investigate the behavior of NMPC for dynamic systems with uncertain parameters. Partial stability for NMPC is established without using terminal costs nor terminal constraints. Numerical examples are provided to illustrate the obtained results.

7.1 Introduction

Nonlinear Model Predictive Control (NMPC) has become the predominant advanced control methodology in recent years. Together with numerical methods, stability analysis for NMPC is widely studied. Stability criteria have been devised for various NMPC scenarios, including infinite horizon control, receding horizon control with zero terminal constraints, Mayne and Michalska [60],[61]; quadratic terminal costs with regional terminal constraints, Chen and Allgöwer [17]; general terminal costs which are control Lyapunov functions, Fontes [29], Jadbabaie and Hauser [41]. See also Findeisen et al. [28], Rawlings and Mayne [71], Grüne and Pannek [36] for comprehensive overviews. Very recent studies concern stability of NMPC without constraints, e.g., Jadbabaie et al. [42], Reble and Allgöwer [72], Grüne [35]. For stability analysis, we consider the following NMPC setup. The dynamic system is given by:

$$\dot{x}(t) = f(t, x(t), u(t)), \quad t \geq 0, \tag{7.1}$$

where $x(t) \in X \subseteq \mathbb{R}^n$, $u(t) \in \mathbb{U} \subseteq \mathbb{R}^{n_u}$, with $0 \in \mathbb{U}$, $0 \in X$ and $f(t, 0, 0) = 0$ for all t. Choose $T > 0$ as a prediction horizon. For each t, x^0, consider the following optimal control problem (OCP)

$$\begin{aligned} \min_{x(\cdot),u(\cdot)} \ & \int_t^{t+T} L(x(\tau), u(\tau))d\tau, \\ \text{s.t.} \ & \dot{x}(\tau) = f(\tau, x(\tau), u(\tau)), \quad \tau \geq t, \\ & x(t) = x^0, \\ & u(\tau) \in \mathbb{U}, \ x(\tau) \in X, \quad \forall \tau \geq t. \end{aligned} \tag{$P_T(t, x^0)$}$$

Control and state bounds can be enforced via the sets $\mathbb{U}$ and X, which we assume to be closed. The function $L : \mathbb{R}^n \times \mathbb{R}^{n_u} \to \mathbb{R}^+$ is often assumed to satisfy

$$L(0,0) = 0,\ L(x,u) \geq \gamma(\|x\|)\ \forall\ x \in X, u \in \mathbb{U}, \tag{7.2}$$

with some $\gamma \in \mathcal{K}_\infty$ to be defined in Section 7.2.

For each t, x^0, suppose that $(P_T(t,x^0))$ has a unique solution with the optimal controls and states $u^*_{T,t,x^0}(\tau)$, $x^*_{T,t,x^0}(\tau)$ and the optimal value $V_T(t,x^0)$. Take a sampling time $T_s < T$ and the grid $t_k = kT_s,\ k = 0,1,2,...$ We consider the following NMPC scheme with sampling.

0. Set $k = 0$.
1. Estimate state $\hat{x}(t_k)$.
2. Solve $(P_T(t_k, \hat{x}(t_k)))$.
3. Apply $u(t) = u^*_{T,t_k,\hat{x}(t_k)}(t)$, $t \in [t_k, t_{k+1})$ to the process.
4. Set $k = k+1$ and go to 1.

For *nominal stability* analysis of NMPC schemes, it is assumed that the estimates $\hat{x}(t_k)$ coincide with the true values. Our setup does not impose any terminal costs or terminal constraints needed in [60, 61, 17, 28]. In fact, it is the state-of-the-art setup which is recently studied in [34, 72, 35]. However, assumption (7.2) which relies on the coerciveness of L with respect to the complete x may be too restrictive, e.g., for economic NMPC, which becomes more and more popular, Diehl et al. [20], Grüne [35]. While there is rich literature on nominal stability, the problem of *partial stability* for NMPC seems to be overlooked. In this chapter, we will use partial stability to weaken assumption (7.2). Moreover, for systems with uncertain parameters and disturbances, the estimates of uncertain parameters may not converge to the true values. It is often the case that, when the states enter a certain region, they provide less and less information to estimate the parameters. One of our goals is to apply the theory of partial stability to the stability analysis for NMPC in such cases. Also when controlling systems under disturbances, we must ensure that the states are (asymptotically) stable provided that the disturbances lie within some suitable region. This is related to the problem of robust control and we will present a treatment of this problem in the framework of partial stability.

The problem of partial stability has been investigated in the theory of differential equations. It was initiated by Lyapunov in 1893 but extensively studied only later in the 1960s. Vorotnikov [84] presents a comprehensive study about this subject. Besides the traditional framework of stability analysis for linear as well as nonlinear systems, Vorotnikov [84] also considers problems of stabilization in connection with optimal control problems.

The chapter is organized as follows. The concepts and main results for partial stability for differential equations are presented in Sections 7.2. We then extend these results to NMPC in Section 7.3, establishing partial stability without terminal costs nor terminal constraints. In Section 7.4 we discuss the behavior of NMPC for dynamic systems with uncertain parameters and provide an illustrative numerical example. We complement the chapter with an appendix for the analytic investigation of the system considered in Section 7.4.

For convenience, we introduce the notation that we use throughout this chapter: By $\boldsymbol{\partial} : \mathbb{R}^n \to \mathbb{R}^m$ we denote a continuous function, which we call the *partiality-mapping* of a vector $x \in \mathbb{R}^n$, e.g., the mapping of x onto its last m components $\boldsymbol{\partial} x = (x_{n-m+1}, x_{n-m+2}, ..., x_n)^T$. In addition, $\mathbb{R}^+$ denotes the set of all nonnegative real numbers.

7.2 Partial stability for differential equations

We consider the following initial value problem (IVP)

$$\begin{aligned} \dot{x}(t) &= \phi(t, x(t)), \\ x(t_0) &= x^0, \end{aligned} \tag{7.3}$$

where $\phi : \mathbb{R}^+ \times D \to \mathbb{R}^n$ is continuous and locally Lipschitz continuous with respect to x. Here, $D = \mathbb{R}^{n-m} \times D_{\boldsymbol{\partial}}$ with an open connected subset $D_{\boldsymbol{\partial}} \subseteq \mathbb{R}^m$ that contains the origin $0 \in \mathbb{R}^m$. The solution of the IVP (7.3) is denoted by $x(t; x^0, t_0)$ or $x(t)$ if the initial conditions are already specified. Without loss of generality, we assume that $x^* = x(t; 0, 0) = 0$ for all $t \geq 0$, i.e., $\phi(t, 0) = 0$.

Definition 7.1 (Partial stability, Vorotnikov [84]). *The solution $x^* = 0$ of (7.3) is said to be $\boldsymbol{\partial}$-stable if for each $t_0 \geq 0, \varepsilon > 0$, there exists $\delta(t_0, \varepsilon) > 0$ such that for all x^0 with $\left\|x^0\right\| < \delta(t_0, \varepsilon)$, we have*

$$\left\|\boldsymbol{\partial} x(t; x^0, t_0)\right\| < \varepsilon \text{ for all } t \geq t_0.$$

Definition 7.2 (Vorotnikov [84]). *The solution $x^* = 0$ of (7.3) is said to be $\boldsymbol{\partial}$- asymptotically stable if it is $\boldsymbol{\partial}$-stable and for each t_0, there exists $\delta_0(t_0) > 0$ such that for any x^0 with $\left\|x^0\right\| < \delta_0(t_0)$, we have*

$$\lim_{t \to \infty} \boldsymbol{\partial} x(t; x^0, t_0) = 0.$$

We use the direct Lyapunov method to investigate the partial stability of (7.3). Consider a continuous function $V : \mathbb{R}^+ \times D \to \mathbb{R}^+$. For $t \geq 0, x \in D$, we set

$$\dot{V}(t, x^0) = \liminf_{h \to 0^+} \frac{V(t + h, x(t + h; x^0, t)) - V(t, x^0)}{h}.$$

The following lemma is essential for the stability proof.

Lemma 7.1 (LaSalle [56]). *If V is continuous in $\mathbb{R}^+ \times D$ and with respect to x locally Lipschitz continuous, then for any $t > t_0$ we have*

$$V(t, x(t; x^0, t_0)) - V(t_0, x^0) \leq \int_{t_0}^{t} \dot{V}(\tau, x(\tau; x^0, t_0)) d\tau. \tag{7.4}$$

We now introduce some important classes of functions which make the investigation of the stability of IVP (7.3) convenient. Define for $R > 0$

$$\mathcal{K}_R = \left\{\alpha : [0, R) \to \mathbb{R}^+, \quad \alpha(0) = 0, \quad \alpha \text{ is continuous, strictly increasing}\right\},$$

with the additional requirement of $\lim_{r \to \infty} \alpha(r) = \infty$ if $R = \infty$. For each $\alpha \in \mathcal{K}_R$, there exists the inverse of α, denoted by $\alpha^{-1} : [0, \bar{\alpha}) \to [0, R)$ with $\bar{\alpha} = \lim_{r \to R^-} \alpha(r) \in \mathbb{R}^+ \cup \{+\infty\}$.

Lemma 7.2. *Assume that there exist a continuous function $V : [0, \infty) \times D \to \mathbb{R}^+$ which is locally Lipschitz continuous with respect to x and satisfying $V(t, 0) = 0$ for all $t \geq 0$, a constant $R > 0$, and functions $\alpha, \gamma \in \mathcal{K}_R$ such that for all $t \geq 0$, $x \in D$, it holds that*

$$\alpha(\|\boldsymbol{\partial} x\|) \leq V(t, x) \quad \textit{and} \quad \dot{V}(t, x) \leq -\gamma(\|\boldsymbol{\partial} x\|).$$

Then the solution $x^* = 0$ of (7.3) is $\boldsymbol{\partial}$-asymptotically stable.

Proof. We first prove that $x^* = 0$ is $\boldsymbol{\partial}$-stable. Given t_0, ε, there exists $\delta(t_0, \varepsilon) < \varepsilon$ such that $V(t_0, x^0) < \alpha(\varepsilon)$ for all $\left\|x^0\right\| \leq \delta(t_0, \varepsilon)$. The ε can be chosen so that $x \in D$ for all $\|\boldsymbol{\partial} x\| \leq \varepsilon$. Take any x^0 with $\left\|x^0\right\| < \delta(t_0, \varepsilon)$. If it was not true that $\left\|\boldsymbol{\partial} x(t; x^0, t_0)\right\| < \varepsilon$ for all $t > t_0$, we could find $t_1 > t_0$ such that for all $t_0 \leq \tau < t_1$

$$\left\|\boldsymbol{\partial} x(\tau; x^0, t_0)\right\| < \varepsilon \quad \text{and} \quad \left\|\boldsymbol{\partial} x(t_1; x^0, t_0)\right\| = \varepsilon. \tag{7.5}$$

Denote $x(t; x^0, t_0)$ by $x(t)$. It follows from (7.4) that

$$\begin{aligned} V(t_1, x(t_1)) - V(t_0, x^0) &\leq \int_{t_0}^{t_1} \dot{V}(\tau, x(\tau)) d\tau \\ &\leq -\int_{t_0}^{t_1} \gamma(\|\boldsymbol{\partial} x(\tau)\|) d\tau \leq 0. \end{aligned} \tag{7.6}$$

Hence

$$\alpha(\|\boldsymbol{\partial} x(t_1)\|) \leq V(t_1, x(t_1)) \leq V(t_0, x^0) < \alpha(\varepsilon).$$

Since α is strictly increasing, $\|\boldsymbol{\partial} x(t_1)\| < \varepsilon$, which is a contradiction to (7.5). Therefore, $\left\|\boldsymbol{\partial} x(t; x^0, t_0)\right\| < \varepsilon$ for all $t > t_0$ and $x^* = 0$ is thus $\boldsymbol{\partial}$-stable.

Choose $r > 0$ so that $\boldsymbol{\partial} x^0 \in D_{\boldsymbol{\partial}}$ for all $\left\|\boldsymbol{\partial} x^0\right\| \leq r$. Because x^* is $\boldsymbol{\partial}$-stable, there exists $\delta_0(t_0) = \delta(t_0, r) > 0$ such that if $\left\|x^0\right\| < \delta_0(t_0)$, then $\left\|\boldsymbol{\partial} x(t; x^0, t_0)\right\| < r$ for all $t \geq t_0$. Since $V(t, x(t)) \geq 0$, we deduce from (7.6) that

$$\int_{t_0}^{t} \gamma(\|\boldsymbol{\partial} x(\tau)\|) d\tau \leq V(t_0, x^0) \text{ for all } t \geq t_0.$$

Consequently, $\lim_{t \to \infty} \gamma(\|\boldsymbol{\partial} x(t)\|) = 0$. Hence,

$$\lim_{t \to \infty} \boldsymbol{\partial} x(t) = 0,$$

which completes the proof. □

The following example, due to LaSalle, see Ballieu and Peiffer [7] shows that the asymptotic stability and partial asymptotic stability are not equivalent.

Example 7.1. *Consider the following system which models the pendulum with friction*

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -(2 + e^t) x_2(t) - x_1(t). \end{cases} \tag{7.7}$$

We set

$$V(t, x) = \frac{1}{2} \left(x_1^2 + x_2^2\right).$$

The derivative of $V(t, x)$ along (7.7) is

$$\dot{V}(t, x) = x_1 x_2 + x_2(-(2 + e^t) x_2 - x_1) = -(2 + e^t) x_2^2 \leq -2 x_2^2.$$

Therefore, (7.7) is stable with respect to the whole states. It is also x_2- asymptotically stable due to Lemma 7.2. However, (7.7) is not asymptotically stable with respect to x_1. In fact, let $a \neq 0$ be an arbitrary real number. The functions

$$x_1(t) = a(1 + e^{-t}), \quad x_2(t) = -ae^{-t}$$

satisfy (7.7) with the initial conditions

$$x_1(0) = 2a, \quad x_2(0) = -a.$$

Clearly

$$\lim_{t\to+\infty} x_1(t) = a.$$

Hence (7.7) is not asymptotically stable. At the end of the chapter we will show that the convergence of $x_1(t)$ and $x_2(t)$ as $t \to \infty$ is of exponential rate.

7.3 Partial stability for NMPC

By applying partial stability, we can weaken assumption (7.2). In fact, we assume that L only satisfies

$$L(0,0) = 0,\ L(x,u) \geq \gamma(\|\boldsymbol{\partial} x\|)\ \forall\ x \in X, u \in \mathbb{U}, \tag{7.8}$$

with some $\gamma \in \mathcal{K}_\infty$. For nonautonomous systems, to ensure that $V_T(t,x^0) \geq \alpha_T(\left\|\boldsymbol{\partial} x^0\right\|)$, we suppose:

(H) There exists an $\alpha_T \in \mathcal{K}_\infty$ such that for any t, x^0 and $u(\tau)$, $x(\tau)$ satisfying (7.1) with $x(t) = x^0$, we have

$$\int_t^{t+T} \gamma(\|\boldsymbol{\partial} x(\tau)\|)d\tau \geq \alpha_T(\left\|\boldsymbol{\partial} x^0\right\|).$$

It is valid for a wide class of systems, as demonstrated in the following proposition.

Proposition 7.1. (H) *holds if f is uniformly bounded and* $\ln(\gamma(r))$ *is uniformly continuous in* $(0,\infty)$.

Proof. Take $M > 0$ such that $\|f(t,x,u)\| < M$ for all t, x, u. Since $\ln(\gamma(r))$ is uniformly continuous for $r > 0$, there exists $\delta > 0$ such that $|\ln(\gamma(r_1)) - \ln(\gamma(r_2))| < \ln(2)$ for $r_1, r_2 > 0, |r_1 - r_2| < \delta$. It follows that $\ln(\gamma(r_1)/\gamma(r_2)) > -\ln(2)$. Hence, $\gamma(r_1)/\gamma(r_2) > \frac{1}{2}$ for $r_1, r_2 > 0, |r_1 - r_2| < \delta$.

Choose $t' = \min\left\{\frac{\delta}{M}, T\right\}$. For any t, x^0 and $t \leq \tau \leq t + t'$, it is easy to see that

$$\left\|\boldsymbol{\partial} x^*_{T,t,x^0}(\tau) - \boldsymbol{\partial} x^0\right\| < M|\tau - t| \leq Mt' \leq \delta.$$

Hence $\gamma(\|\boldsymbol{\partial} x^*_{T,t,x^0}(\tau)\|) \geq \frac{1}{2}\gamma(\|\boldsymbol{\partial} x^0\|)$ for all $t \leq \tau \leq t + t'$. Consequently

$$\begin{aligned} V_T(t,x^0) &= \int_t^{t+T} L(x^*_{T,t,x^0}(\tau), u^*_{T,t,x^0}(\tau))d\tau \\ &\geq \int_t^{t+T} \gamma\left(\left\|\boldsymbol{\partial} x^*_{T,t,x^0}(\tau)\right\|\right) d\tau \\ &\geq \int_t^{t+t'} \gamma\left(\left\|\boldsymbol{\partial} x^*_{T,t,x^0}(\tau)\right\|\right) d\tau \geq \frac{1}{2}t'\gamma(\left\|\boldsymbol{\partial} x^0\right\|). \end{aligned}$$

Taking $\alpha_T(r) = \frac{1}{2}t'\gamma(r)$ yields the desired conclusion. □

Lemma 7.3. *Suppose that $0 \leq T_1 \leq T_2$. Then for any $t \geq 0, x^0 \in X$ we have $V_{T_2}(t, x^0) \geq V_{T_1}(t, x^0)$.*

Proof. The control $u^*_{T_2,t,x^0}(\tau)$ for $\tau \in [t, t+T_1]$ is feasible for $(P_{T_1}(t, x^0))$. Hence

$$\begin{aligned} V_{T_2}(t, x^0) &= \int_t^{t+T_2} L(x^*_{T_2,t,x^0}(\tau), u^*_{T_2,t,x^0}(\tau))d\tau \\ &\geq \int_t^{t+T_1} L(x^*_{T_2,t,x^0}(\tau), u^*_{T_2,t,x^0}(\tau))d\tau \geq V_{T_1}(t, x^0). \end{aligned}$$

□

Let us denote by $x(\tau)$, $u(\tau)$ the solution produced by the NMPC scheme with the initial states $x(0)$. We define $\tilde{L}_k(\tau) = L(x^*_{T,t_k,x(t_k)}(\tau), u^*_{T,t_k,x(t_k)}(\tau))$ for $\tau \in [t_k, t_k + T]$ and $\tilde{L} : [0, \infty) \to \mathbb{R}^+$ piecewise by

$$\tilde{L}(\tau) = L(x^*_{T,t_k,x(t_k)}(\tau), u^*_{T,t_k,x(t_k)}(\tau)) \text{ for } \tau \in [t_k, t_{k+1}).$$

Here is the main result of the chapter:

Theorem 7.1. *If* (H) *is satisfied, V_T is continuous and there exists $\beta \in (0, 1]$ such that for $k = 0, 1, 2, \ldots$, it holds that*

$$V_T(t_{k+1}, x(t_{k+1})) \leq V_T(t_k, x(t_k)) - \beta \int_{t_k}^{t_{k+1}} \tilde{L}(\tau)d\tau. \tag{7.9}$$

Then the closed-loop system produced by the NMPC scheme is $\boldsymbol{\partial}$-asymptotically stable.

Proof. We define the function

$$V(t) = V_T(t_k, x(t_k)) - \beta \int_{t_k}^{t} \tilde{L}(\tau)d\tau$$

piecewise for $t \in [t_k, t_{k+1})$. By the Bellman dynamic programming principle (DP),

$$V_{T-t+t_k}(t, x(t)) = V_T(t_k, x(t_k)) - \int_{t_k}^{t} \tilde{L}(\tau)d\tau.$$

Hence $V(t) \geq V_{T-t+t_k}(t, x(t))$. Since $T - t + t_k \geq T - t_{k+1} + t_k = T - T_s$, Lemma 7.3 gives $V_{T-t+t_k}(t, x(t)) \geq V_{T-T_s}(t, x(t))$. We then deduce from (7.8) and (H) that

$$V(t) \geq V_{T-T_s}(t, x(t)) \geq \alpha_{T-T_s}(\|\boldsymbol{\partial} x(t)\|).$$

Suppose that $\tau_0 < \tau_1$ and $\ell \leq k$ such that $\tau_0 \in [t_\ell, t_{\ell+1})$, $\tau_1 \in [t_k, t_{k+1})$. We have by definition

$$\begin{aligned} V(\tau_1) &= V_T(t_k, x(t_k)) - \beta \int_{t_k}^{\tau_1} \tilde{L}(\tau)d\tau, \\ V(\tau_0) &= V_T(t_\ell, x(t_\ell)) - \beta \int_{t_\ell}^{\tau_0} \tilde{L}(\tau)d\tau. \end{aligned}$$

Therefore

$$V(\tau_1) - V(\tau_0) = V_T(t_k, x(t_k)) - V_T(t_\ell, x(t_\ell)) + \beta \int_{t_\ell}^{\tau_0} \tilde{L}(\tau)d\tau - \beta \int_{t_k}^{\tau_1} \tilde{L}(\tau)d\tau$$

$$
\begin{aligned}
&= \sum_{i=\ell}^{k-1} [V_T(t_{i+1}, x(t_{i+1})) - V_T(t_i, x(t_i))] \\
&\quad + \beta \int_{t_\ell}^{\tau_0} \tilde{L}(\tau) d\tau - \beta \int_{t_k}^{\tau_1} \tilde{L}(\tau) d\tau \\
&\leq -\beta \sum_{i=\ell}^{k-1} \int_{t_i}^{t_{i+1}} \tilde{L}(\tau) d\tau + \beta \int_{t_\ell}^{\tau_0} \tilde{L}(\tau) d\tau - \beta \int_{t_k}^{\tau_1} \tilde{L}(\tau) d\tau \\
&= -\beta \int_{\tau_0}^{\tau_1} \tilde{L}(\tau) d\tau.
\end{aligned}
$$

Together with (7.8), it gives

$$
V(\tau_1) - V(\tau_0) \leq -\int_{\tau_0}^{\tau_1} \beta\gamma(\|\boldsymbol{\partial} x(\tau)\|) d\tau \leq 0. \tag{7.10}
$$

Now we prove that $x^* = 0$ is $\boldsymbol{\partial}$-stable. Given $\tau_0, \varepsilon > 0$, one can choose ℓ and $\delta_\ell < \varepsilon$ so that $t_\ell \leq \tau_0 < t_{\ell+1}$ and

$$
V_T(t_{\ell+1}, \bar{x}^0) < \alpha_{T-T_s}(\varepsilon), \quad \forall \left\|\bar{x}^0\right\| \leq \delta_\ell. \tag{7.11}
$$

This is because $V_T(t, 0) = 0$ and V_T is continuous. Since $x(t; \tau_0, x^0)$ is continuous in (t, x^0), there exists $\delta > 0$ such that

$$
\|x(t)\| \leq \delta_\ell, \quad \forall \left\|x^0\right\| \leq \delta, \ \tau_0 \leq t \leq t_{\ell+1}
$$

where $x(t) = x(t; \tau_0, x^0)$. For any $\left\|x^0\right\| \leq \delta$, by the choice of δ, $\|\boldsymbol{\partial} x(t)\| \leq \|x(t)\| \leq \delta_\ell < \varepsilon$ on $[\tau_0, t_{\ell+1}]$. For $t > t_{\ell+1}$, we have

$$
\begin{aligned}
\alpha_{T-T_s}(\|\boldsymbol{\partial} x(t)\|) &\leq V(t) \leq V(t_{\ell+1})) \\
&= V_T(t_{\ell+1}, x(t_{\ell+1})) < \alpha_{T-T_s}(\varepsilon).
\end{aligned}
$$

in view of (7.10) and (7.11). Because α_{T-T_s} is strictly increasing, this implies that $\|\boldsymbol{\partial} x(t)\| < \varepsilon$ for all $t \geq \tau_0$. Furthermore, since $V(t) \geq 0$, we deduce from (7.10) that

$$
\int_{\tau_0}^{t} \beta\gamma(\|\boldsymbol{\partial} x(\tau)\|) d\tau \leq V(\tau_0), \ \forall t \geq \tau_0.
$$

Consequently, $\lim_{t\to\infty} \gamma(\|\boldsymbol{\partial} x(t)\|) = 0$. Hence

$$
\lim_{t\to\infty} \boldsymbol{\partial} x(t) = 0,
$$

which completes the proof. □

The following lemma provides a means to verify (7.9), cf. [72] where continuity of B is redundantly assumed.

Lemma 7.4. *If there exists a non-decreasing, bounded function $B : \mathbb{R}^+ \to \mathbb{R}^+$ such that*

$$
V_T(t, x^0) \leq B(T) L^*(x^0), \tag{7.12}
$$

for any $x^0 \in X$, $t, T \geq 0$, where $L^*(x^0) = \min_{u\in\mathbb{U}} L(x^0, u)$, then with a suitable choice of T, condition (7.9) is fulfilled.

Proof. Take $k \geq 0$. Let $h \in [T_s, T]$ and define the control $\hat{u}_h(t)$ on $[t_{k+1}, t_{k+1} + T]$ as follows,

$$\hat{u}_h(t) = \begin{cases} u^*_{T,t_k,x(t_k)}(t) & \text{if } t \in [t_{k+1}, t_k + h), \\ u^*_{T,t_k+h,x_{kh}}(t) & \text{if } t \in [t_k + h, t_{k+1} + T]. \end{cases}$$

where $x_{kh} = x^*_{T,t_k,x(t_k)}(t_k + h)$. Also denote by $\hat{x}_h(t)$ the corresponding states. We have $\hat{u}_h(t) \in \mathbb{U}$. On $[t_{k+1}, t_k + h]$, $\hat{x}_h(t) = x^*_{T,t_k,x(t_k)}(t) \in X$ and on $[t_k + h, t_{k+1} + T]$, $\hat{x}_h(t) = x^*_{T,t_k+h,x_{kh}}(t) \in X$. Hence $\hat{u}_h(t)$ is feasible. For $t \in [t_k + h, t_{k+1} + T]$, set $\hat{L}(t) = L(x^*_{T,t_k+h,x_{kh}}(t), u^*_{T,t_k+h,x_{kh}}(t))$. If we evaluate the objective function at $\hat{u}_h(t)$, $\hat{x}_h(t)$, then due to the optimality of $V_T(t_{k+1}, x(t_{k+1}))$, it holds that,

$$\begin{aligned} V_T(t_{k+1}, x(t_{k+1})) &\leq \int_{t_{k+1}}^{t_k+h} \tilde{L}_k(\tau)d\tau + \int_{t_k+h}^{t_{k+1}+T} \hat{L}(\tau)d\tau \\ &= \int_{t_{k+1}}^{t_k+h} \tilde{L}_k(\tau)d\tau + V_{T-h+T_s}(t_k + h, x^*_k(t_k + h)) \\ &\leq \int_{t_{k+1}}^{t_k+h} \tilde{L}_k(\tau)d\tau + B(T - h + T_s)L^*(x^*_k(t_k + h)), \end{aligned}$$

by virtue of (7.12) where $x^*_k(t) = x^*_{T,t_k,x(t_k)}(t)$. As a result

$$\begin{aligned} V_T(t_{k+1}, x(t_{k+1})) &\leq \int_{t_{k+1}}^{t_k+T} \tilde{L}_k(\tau)d\tau + B(T) \min_{h\in[T_s,T]} L^*(x^*_k(t_k + h)) \\ &\leq \int_{t_{k+1}}^{t_k+T} \tilde{L}_k(\tau)d\tau + \frac{B(T)}{T - T_s} \int_{t_{k+1}}^{t_k+T} L^*(x^*_k(\tau))d\tau. \end{aligned} \tag{7.13}$$

The last inequality is a consequence of the mean value theorem for integrals. Now let $h \in [0, T_s]$. By the DP

$$V_{T-h}(t_k + h, x(t_k + h)) = \int_{t_k+h}^{t_k+T} \tilde{L}_k(\tau)d\tau.$$

Therefore

$$\begin{aligned} \int_{t_{k+1}}^{t_k+T} \tilde{L}_k(\tau)d\tau &\leq \int_{t_k+h}^{t_k+T} \tilde{L}_k(\tau)d\tau = V_{T-h}(t_k + h, x(t_k + h)) \\ &\leq B(T)L^*(x(t_k + h)). \end{aligned}$$

Again by the mean value theorem for integrals

$$\begin{aligned} \int_{t_{k+1}}^{t_k+T} \tilde{L}_k(\tau)d\tau &\leq B(T) \min_{h\in[0,T_s]} L^*(x(t_k + h)) \\ &\leq B(T)\frac{1}{T_s} \int_{t_k}^{t_{k+1}} L^*(x(\tau))d\tau. \end{aligned} \tag{7.14}$$

Note that $L^*(x^*_k(\tau)) \leq \tilde{L}_k(\tau)$. From (7.13)-(7.14) follows

$$V_T(t_{k+1}, x(t_{k+1})) \leq \int_{t_{k+1}}^{t_k+T} \tilde{L}_k(\tau)d\tau + \frac{B(T)^2}{T_s(T - T_s)} \int_{t_k}^{t_{k+1}} \tilde{L}(\tau)d\tau.$$

On the other hand

$$\begin{aligned} V_T(t_k, x(t_k)) &= \int_{t_k}^{t_k+T} \tilde{L}_k(\tau) d\tau \\ &= \int_{t_{k+1}}^{t_k+T} \tilde{L}_k(\tau) d\tau + \int_{t_k}^{t_{k+1}} \tilde{L}(\tau) d\tau. \end{aligned}$$

Hence,

$$V(t_{k+1}, x(t_{k+1})) - V(t_k, x(t_k)) \leq -\left(1 - \frac{B(T)^2}{T_s(T-T_s)}\right) \int_{t_k}^{t_{k+1}} \tilde{L}(\tau) d\tau.$$

Recall that $B(T)$ is bounded. Fixing T_s, we can choose $T > T_s$ so that $0 < \beta = 1 - \frac{B(T)^2}{T_s(T-T_s)} < 1$. We then obtain

$$V_T(t_{k+1}, x(t_{k+1})) - V_T(t_k, x(t_k)) \leq -\beta \int_{t_k}^{t_{k+1}} \tilde{L}(\tau) d\tau.$$

This completes the proof. □

7.4 NMPC with uncertain parameters

For some systems with unknown parameters, we can use an NMPC strategy together with moving horizon estimation (MHE) to estimate parameters and states online. It is often the case that, although the parameters do not converge to the true values, asymptotic stability of the states can still be observed. This can be explained from the view-point of partial stability. In fact, the parameters can be considered as extra states of an augmented dynamic system by setting $\dot{p}(t) = 0$. It happens that when these extra states lie in some suitable region, the states we are interested in enter the *region of attraction*.

For an example, we consider the following system

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), \quad t \in [0; t_f], \\ \dot{x}_2(t) &= -(2 + e^t)x_2(t) - x_1(t) + p(t)x_2^2(t) + u(t), \\ \dot{p}(t) &= 0, \end{aligned} \tag{7.15}$$

where $u(t)$ is a control and $p(t)$ is a constant parameter. Without $p(t)$ and $u(t)$, (7.15) becomes linear and Ballieu and Peiffer [7] showed that it is $\boldsymbol{\partial}$-asymptotically stable with $\boldsymbol{\partial}x = x_2$ but not asymptotically stable. In contrast to that, the presence of $px_2^2(t)$ and $u(t)$ makes the analysis of (7.15) much more difficult.

Proposition 7.2. *Suppose that p is given and $u(t)$ is bounded. For any solution $x(t)$ of (7.15) with $\|x(0)\|$ sufficiently small, $x_2(t)$ converges to 0 exponentially and $x_1(t)$ tends to a certain limit as $t \to \infty$. As a consequence, (7.15) is $\boldsymbol{\partial}$-asymptotically stable for $\boldsymbol{\partial}x = x_2$.*

We give the proof of Proposition 7.2 at the end of this chapter. A natural question arises: How will the system behave if we try to control the states to the origin? To

answer it, we consider the following OCP

$$\begin{aligned} \min_{u(\cdot),x(\cdot)} \quad & \int_0^{t_f} \left(w_1 x_1^2(t) + w_2 x_2^2(t) + u^2(t)\right) dt \\ \text{s.t.} \quad & (7.15) \text{ holds}, \\ & -5 \le u(t) \le 5. \end{aligned} \tag{7.16}$$

Here $w = (w_1; w_2) > 0$ are adjustable weights.

We first simulate the system with the parameter $p \equiv 1$. The initial values are $x(0) = (1, 2)^T$. The final time is $t_f = 10$. Figure 7.1 shows the behavior of (7.15) under varying weights. With increasing weights on x_1, the control gradually forces x_1 to the origin. At the same time, x_2 automatically tends to 0 regardless whether we want to control it or not, see Figure 7.1 with $w = (10; 0)$. This is due to the inherent asymptotic stability of x_2.

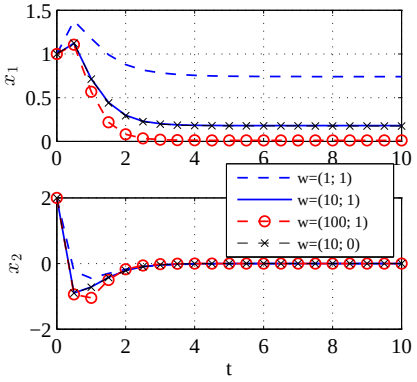

Figure 7.1: With increasing weights on x_1, the control better steers x_1 towards 0. Even when we do not put weight on x_2, it still converges to 0.

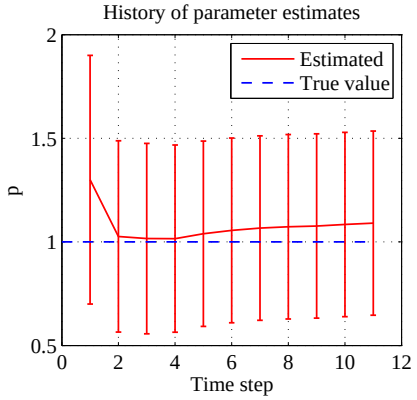

Figure 7.2: Parameter estimates are plotted with symmetric error bars of two times standard deviation. As x_2 tends to 0 rapidly, less information is gained to estimate p leading to less improvement on the accuracy of the estimates.

Now consider the case when the parameter p is unknown. We employ NMPC with moving horizon estimation (MHE) following the setup: $w = (1; 1)$; the initial guess is

$p_0 = 1.3$ with a prior variance of $P_0 = 0.04$; prediction horizon $T = 10$ with the sampling time $T_s = 0.5$. We carry out 10 NMPC steps. At each step, $N_c = 2$ control elements are applied to the system.

Numerical simulation indicates that, because of the rapid convergence of x_2 to 0, the parameter has less and less effect to the system (Figure 7.2). It causes severe inaccuracy in the estimates, however, without aggravating the overall performance. The optimal cost computed in the perfect case when p is fixed at the true value is 10.0975 while the cost with NMPC-MHE is 10.1546, i.e., only 0.5% worse.

Summary. We prove a stability result for NMPC based on partial stability, which requires less restrictive assumptions than comparable conventional stability proofs. We then investigate the stability of NMPC with MHE for the control of dynamic systems with uncertain parameters. We observe that, for some specific systems, due to the inherent partial stability, little information for estimating some parameters is gained. The estimates are poor but the overall performance is almost unaffected.

Proof of Proposition 7.2

We now give a detailed investigation of system (7.15). First, we consider its linearization around 0

$$\begin{cases} \dot{z}_1(t) = z_2(t), \quad t \in [0; t_f] \\ \dot{z}_2(t) = -(2 + e^t) z_2(t) - z_1(t) + u(t). \end{cases} \tag{7.17}$$

Without $u(t)$ it reads as

$$\dot{\bar{z}}(t) = \begin{pmatrix} 0 & 1 \\ -1 & (-2 - e^t) \end{pmatrix} \bar{z}(t) \tag{7.18}$$

One solution of (7.18) is

$$\bar{z}_1(t) = 1 + e^{-t}; \quad \bar{z}_2(t) = -e^{-t}.$$

Using the Liouville-Ostrogradski formula, we can find a solution $y(t) = (y_1(t), y_2(t))$ so that $\bar{z}(t)$ and $y(t)$ form basic solutions of (7.18). Those are

$$\begin{aligned} y_1(t) &= (1 + e^{-t}) \int_0^t \frac{\exp(-e^\tau)}{(1 + e^\tau)^2} d\tau; \\ y_2(t) &= \frac{\exp(-e^t)}{e^t(1 + e^t)} - e^{-t} \int_0^t \frac{\exp(-e^\tau)}{(1 + e^\tau)^2} d\tau. \end{aligned}$$

Obviously for these basic solutions, $\bar{z}_1(t), y_1(t)$ are convergent as $t \to +\infty$ and there exists a constant $C > 0$ such that

$$|\bar{z}_2(t)|, |y_2(t)| \le Ce^{-t}. \tag{7.19}$$

The inverse of the corresponding fundamental matrix is

$$\Phi^{-1}(t) = \exp(e^t + 2t) \begin{pmatrix} y_2(t) & -y_1(t) \\ -\bar{z}_2(t) & \bar{z}_1(t) \end{pmatrix}.$$

In the following, we denote several constants by C, although the actual values may be different.

Proposition 7.3. *For any solution $z(t)$ of (7.17), there exist $C > 0, 0 < \beta < 1$, such that $|z_2(t)| \leq Ce^{-\beta t}$ for all $t \geq 0$. Furthermore, $z_1(t)$ converges as $t \to \infty$.*

Proof. The solution $z(t)$ admits a representation via the fundamental matrix

$$z(t) = \Phi(t)\Phi^{-1}(0)z(0) + \Phi(t)\int_0^t \Phi^{-1}(\tau)\begin{pmatrix}0\\ u(\tau)\end{pmatrix}d\tau.$$

We first prove that $z_2(t) \to 0$ as $t \to +\infty$. In fact

$$\begin{aligned} z_2(t) =&[\bar{z}_2(t) \quad y_2(t)]\Phi^{-1}(0)z(0) - \bar{z}_2(t)\int_0^t y_1(\tau)u(\tau)\exp(e^\tau + 2\tau)d\tau \\ &+ y_2(t)\int_0^t \bar{x}_1(\tau)u(\tau)\exp(e^\tau + 2\tau)d\tau. \end{aligned} \tag{7.20}$$

The first term of (7.20) tends to 0 as $t \to \infty$ because of (7.19). The sum of the other terms is

$$\begin{aligned} &e^{-t}\int_0^t y_1(\tau)u(\tau)\exp(e^\tau + 2\tau)d\tau \\ &- e^{-t}\int_0^t \bar{z}_1(\tau)u(\tau)\exp(e^\tau + 2\tau)d\tau\int_0^t \frac{\exp(-e^\tau)}{(1+e^\tau)^2}d\tau \\ &+ \frac{\exp(-e^t)}{e^t(1+e^t)}\int_0^t \bar{z}_1(\tau)u(\tau)\exp(e^\tau + 2\tau)d\tau. \end{aligned} \tag{7.21}$$

The third term of (7.21) which is positive has the limit

$$\lim_{t\to\infty}\frac{\int_0^t \bar{z}_1(\tau)u(\tau)\exp(e^\tau + 2\tau)d\tau}{\exp(e^t + 2t)} = \lim_{t\to\infty}\frac{\bar{z}_1(t)u(t)\exp(e^t + 2t)}{(2+e^t)\exp(e^t + 2t)} = 0,$$

by virtue of the L'Hôpital rule. Similarly, the sum of the first two terms of (7.21) also tends to 0 as $t \to \infty$. Thus $\lim_{t\to\infty} z_2(t) = 0$. As a consequence, $z_2(t)$ is bounded. There exists $B > 0$ such that $|z_2(t)| \leq B$ for $t \geq 0$. Therefore

$$|z_1(t)| \leq |z_1(0)| + Bt.$$

It follows from the second equation in (7.17) that

$$z_2(t) = e^{(-e^t - 2t)}\left(C - \int_0^t (z_1(\tau) + u(\tau))e^{(e^\tau + 2\tau)}d\tau\right).$$

For any $0 < \beta < 1$, applying the L'Hôpital rule, we have

$$\begin{aligned} \lim_{t\to\infty} e^{\beta t}z_2(t) &= \lim_{t\to\infty}\frac{C - \int_0^t (z_1(\tau) + u(\tau))\exp(e^\tau + 2\tau)d\tau}{\exp(e^t + (2-\beta)t)} \\ &= \lim_{t\to\infty}\frac{-(z_1(t) + u(t))\exp(e^t + 2t)}{(e^t + 2 - \beta)\exp(e^t + (2-\beta)t)} \\ &= \lim_{t\to\infty}\frac{-(z_1(t) + u(t))}{\exp[(1-\beta)t]} = 0. \end{aligned}$$

It follows immediately that there exists

$$\lim_{t\to\infty} z_1(t) = z_1(0) + \int_0^\infty z_2(\tau)d\tau.$$

□

We are now ready to prove Proposition 7.2.

Proof. Set

$$K(t,s) = -\bar{z}_2(t)e^{(e^s+2s)}y_1(s) + y_2(t)e^{(e^s+2s)}\bar{z}_1(s).$$

Suppose $(x_1(t), x_2(t))$ is a solution of (7.15). Then we have the representation

$$x_2(t) = [\bar{z}_2(t) \quad y_2(t)]\Phi^{-1}(0)x_0 + \int_0^t K(t,s)(px_2^2(s) + u(s))ds. \tag{7.22}$$

In view of Proposition 7.3,

$$z_2(t) = [\bar{z}_2(t) \quad y_2(t)]\Phi^{-1}(0)z(0) + \int_0^t K(t,s)u(s)ds.$$

Therefore

$$\left|\int_0^t K(t,s)u(s)ds\right| \le Ce^{-\beta t}. \tag{7.23}$$

We show that there exist $C > 0, 0 < \alpha \le \beta$ such that

$$|K(t,s)| \le Ce^{-\alpha(t-s)} \text{ for all } 0 \le s \le t. \tag{7.24}$$

In fact,

$$\begin{aligned} K(t,s) &= e^{-t}\exp(e^s+2s)(1+e^{-s})\int_0^s \frac{\exp(-e^\tau)}{(1+e^\tau)^2}d\tau \\ &+\Big[\frac{\exp(-e^t)}{e^t(1+e^t)} - \frac{1}{e^t}\int_0^t \frac{\exp(-e^\tau)}{(1+e^\tau)^2}d\tau\Big]\exp(e^s+2s)(1+e^{-s}) \\ &= e^{-t}\exp(e^s+2s)(1+e^{-s})\Big[\frac{\exp(-e^t)}{1+e^t} - \int_s^t \frac{\exp(-e^\tau)}{(1+e^\tau)^2}d\tau\Big]. \end{aligned}$$

Choose $0 < \alpha < 1 - e^{-1}$, $\alpha \le \beta$. We prove that

$$|K(t,s)| \le 2e^{-\alpha(t-s)} \text{ for all } 0 \le s \le t.$$

This is equivalent to

$$\begin{aligned} \exp(e^s+2s)(1+e^{-s})\Big|\frac{\exp(-e^t)}{1+e^t} - \int_s^t \frac{\exp(-e^\tau)}{(1+e^\tau)^2}d\tau\Big| \\ \le 2e^{(1-\alpha)t+\alpha s} \quad \text{for all } 0 \le s \le t. \end{aligned}$$

which can be proved by elementary differentiation, considering s fixed and t varying.

We are now in a position to prove the convergence of $x_2(t)$ to 0 as $t \to \infty$. From (7.19), (7.22)-(7.24)

$$|x_2(t)| \le Ce^{-\alpha t} + C\int_0^t e^{-\alpha(t-s)}|x_2(s)|^2 ds. \tag{7.25}$$

Choose $\delta > 0$ such that $\delta C < \alpha$. For any $|x_2(0)| < \delta$, there exists $t_1 > 0$ such that

$$|x_2(t)| \le \delta \text{ for all } 0 \le t \le t_1.$$

It follows from (7.25) that

$$|x_2(t)|e^{\alpha t} \leq C + \int_0^t C\delta e^{\alpha s}|x_2(s)|ds, \text{ for all } 0 \leq t \leq t_1.$$

The Gronwall-Bellman inequality gives us

$$|x_2(t)|e^{\alpha t} \leq C \exp\,(\int_0^t C\delta ds)$$

or

$$|x_2(t)| \leq Ce^{-(\alpha - C\delta)t}, \text{ for all } 0 \leq t \leq t_1.$$

As a consequence, by choosing $x_2(0)$ sufficiently small, one can choose t_1 to be ∞. The last inequality then holds for all $t \geq 0$. Thus, $x_2(t)$ converges to 0 exponentially. The claim that $x_1(t)$ converges to a finite limit exponentially as $t \to \infty$ is proved as in Proposition 7.3 . □

Chapter 8

Applications

In this chapter we present a collection of examples to illustrate the performance of dual NMPC as well as various aspects of controlling uncertain systems. The rocket car example and the moon lander example are classics in optimal control theory that admit analytic solutions. They clearly exhibit the two extremes of the interplay between the performance control task and the information gain task. That is for a parameter, the two tasks are completely contradicting and for the other parameter, the two tasks work identically. The more sophisticated example of a batch bio-reactor displays unstable behavior of nominal control in the estimation procedure, in which nominal control can lead to divergence of parameter estimates. Finally, in the example of a tractor passing the corner, measurement noise is observed to provide excitation, helping to estimate important parameters. This leads to a satisfactory performance of nominal control.

8.1 A rocket car: Conflict and agreement between information gain and performance control

Consider an object that can accelerate and brake. We aim to steer it on a straight line from point A to point B in minimal time. The model can be described as

$$\begin{cases} \dot{x}_1(t) = x_2(t), \\ \dot{x}_2(t) = a u_1(t) - b u_2(t), \end{cases} \tag{8.1}$$

where x_1 is the position and x_2 is the velocity. The controls include the acceleration u_1 and the brake u_2 and are subject to constraints

$$0 \le u_1 \le 2, \quad 0 \le u_2 \le 2. \tag{8.2}$$

There are two uncertain parameters a and b which represent the acceleration and braking coefficients, respectively. A final time T is called feasible if there exist piecewise continuous functions $u_1(t), u_2(t) : [0, T] \to [0, 2]$ such that

$$x(0) = 0,\ y(0) = 0; \quad x(T) = 1,\ y(T) = 0.$$

The goal is to solve in the set of all feasible final times and corresponding control functions the problem

$$\min_{u,T} T.$$

Analytic solution. Using the Pontryagin Minimum Principle, one can show that optimal controls are of bang-bang type, see Example 2.2. Denoting by T^* the optimal final time and t_c^* the time at which the object stops accelerating and starts braking, i.e.,

$$\begin{aligned} u_1(t) &= 2, \quad u_2(t) = 0 \text{ on } [0, t_c^*], \\ u_1(t) &= 0, \quad u_2(t) = 2 \text{ on } (t_c^*, T^*], \end{aligned}$$

we can find that

$$T^* = \sqrt{\frac{a+b}{ab}}, \quad t_c^* = \frac{bT^*}{a+b}.$$

For $a = 2, b = 2$, we have $T^* = 1$ and $t_c^* = 0.5$.

Transformation into an OCP with fixed final time. Suppose that q is an optimal final time, which is unknown. Set

$$\tau = t/q, \quad \tau \in [0, 1].$$

Introducing an auxiliary state $x_3(\tau)$, we formulate the original problem as the following OCP

$$\min_{q, u_1, u_2} x_3(1), \tag{8.3}$$

subject to

$$\begin{aligned} \dot{x}_1(\tau) &= q x_2(\tau), \\ \dot{x}_2(\tau) &= q(u_1(\tau) - b u_2(\tau)), \\ \dot{x}_3(\tau) &= q, \qquad \tau \in [0, 1] \end{aligned}$$

with conditions

$$\begin{gathered} x_1(0) = 0,\ x_2(0) = 0, \quad x_1(1) = 1,\ x_2(1) = 0; \\ 0 \le u_1(\tau), u_2(\tau) \le 2, \quad \tau \in [0, 1]; \\ q > 0. \end{gathered}$$

Numerical results. For numerical computation, we first use a conventional transformation to transform the OCP at hand into OCP (8.3). Suppose that the true parameters are $a^* = 2, b^* = 2$. The perfect controls are obtained by solving a single OCP with the true values of the parameters and states. Perfect optimal controls are of bang-bang type, see Figure 8.1. The optimal time is computed to be $T^* = 1$ and the switching time is $t_c^* = 0.5$, which verify the analytic optimal solution given earlier.

Consider now the case of unknown a, b. Our initial guesses are $a_0 = 3, b_0 = 3$. Their a priori variances are both 1. We set up NMPC as follows. We discretize the interval $[0, 1]$ by $N = 36$ grid points and use a control horizon of $N_c = 3$, i.e., 12 NMPC iterations are carried out in each run. For the penalizing dual NMPC presented Section 6.4.1, we use a dual horizon $N_d = 6$ and a weight $\alpha = 100$. For the measurement function we choose $\eta(x) = (x_1, x_2)^T$ with noise variance 0.001.

Figures 8.1–8.2 illustrate nominal NMPC solutions. With the braking coefficient $b = 3$, the switching time t_c is bigger than t_c^*. During acceleration, $u_2 \equiv 0$, causing non-identifiability of b. This means that no improvement in estimating b is made during this period. After that, the braking is active and the estimate of b gets more precise.

However, because the object is now too close to B and the velocity is high, the object fails to reach the target exactly. Violation of the end-point constraints is as high as 30%.

Dual NMPC, see Figures 8.3–8.2 does a better job. It first tells the controller to exercise maximal acceleration and a little braking. The parameter estimates are good enough at the optimal switching time leading to acceptable feasibility at the end-point. The final constraints are satisfied at 99.4%. Optimality with respect to the case without uncertainty is certainly lost to some degree due the interweaving of acceleration and braking at the beginning. The final time computed by dual control is $T_D = 1.22$ (greater than $T^* = 1$).

Remark 8.1. *Through the rocket car example, we can realize the behavior of performance control and information gain in estimating particular parameters. At the beginning, the performance control task tries to accelerate maximally and it is exactly what the information gain task does for estimating a. However by using no braking, the performance control task gives no information for estimating b. Thus is completely against the information gain task.*

Remark 8.2. *Our setting of solving OCPs with free final time generally leads a nonuniform time grid. This is elucidated as follows. At the ith NMPC step, we solve OCP* (8.3) *on the horizon $N_i < N$. If the optimal value is q_i^*, then the grid size at the moment is q_i^*/N_i, which changes as i varies. Moreover, due to numerical approximation, the computed controls may not be truly bang-bang although the analytic ones are.*

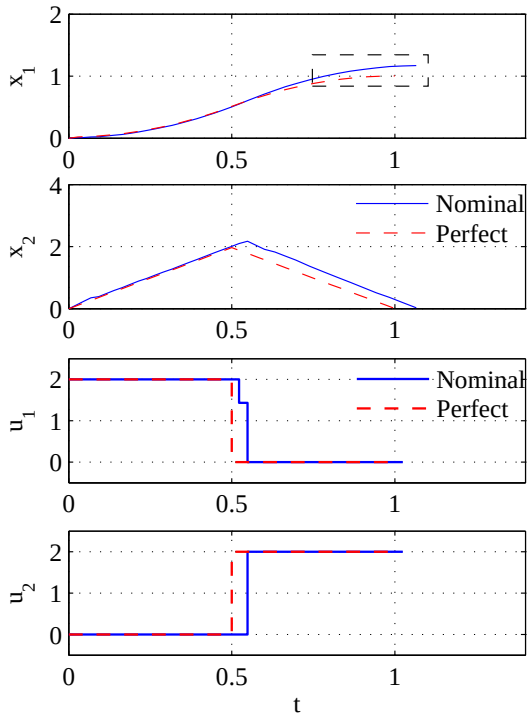

Figure 8.1: Optimal controls (two plots below) for the rocket car are of bang-bang type. Compared with the perfect case, nominal NMPC yields longer accelerating duration, leading to intolerable violation of constraints (two plots above)

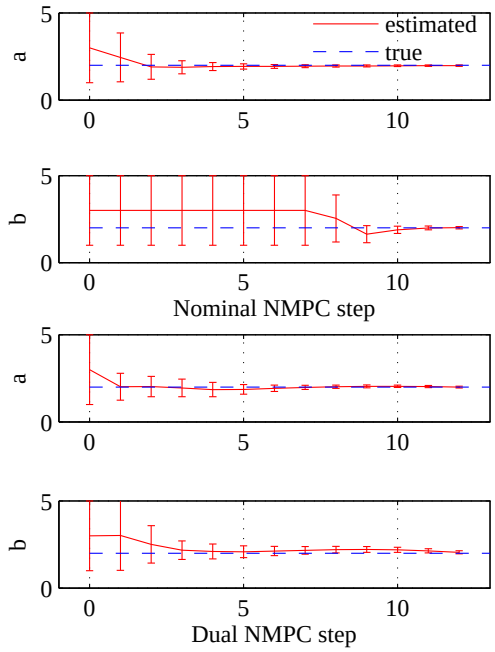

Figure 8.2: Parameter estimates are plotted with error bars which are symmetric and 2 times standard deviation long. Nominal NMPC gives no improvement in the estimates of b at the beginning while dual NMPC actively reduces their variances. At the same time, the estimates of a are fine in both cases.

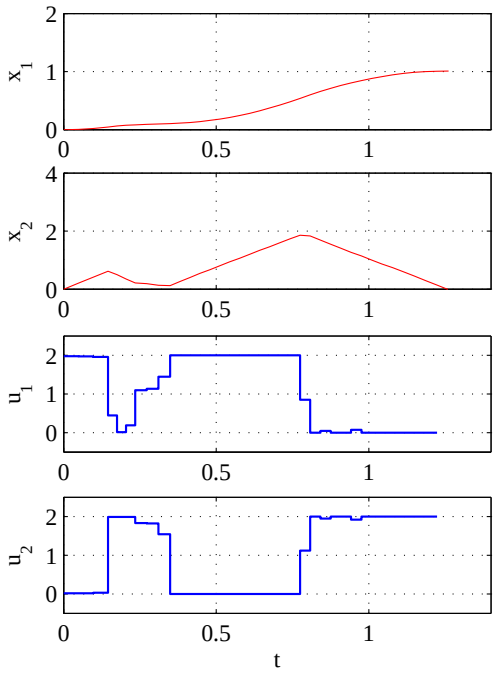

Figure 8.3: Dual NMPC for the rocket car problem uses braking earlier in order to estimate the braking coefficient on time, helping to ensure feasibility.

8.2 A moon lander problem

The model reads as (see Evans [25])

$$\begin{aligned} \dot{h}(t) &= v(t), \\ \dot{v}(t) &= -g + \frac{u(t)}{m(t)}, \\ \dot{m}(t) &= -pu(t). \end{aligned} \tag{8.4}$$

The involved quantities are explained in Table 8.1. The mass is never negative and the

Notation	Description	Unit
h	Altitude (distance of the lander to the moon surface)	m
v	Velocity	$m^{-1}s$
m	Mass of the lander (decreases as fuel is burned)	kg
u	Thrust of the rocket	N
p	Coefficient of the thrust	$m^{-1}s$
g	Gravitation of the moon, known to be 1.625	ms^{-2}

Table 8.1: Description of variables in the moon lander problem

lander must not penetrate the ground. Therefore we impose the constraints $m(t) \geq 0,\ h(t) \geq 0$. Also the thrust is bounded

$$0 \leq u(t) \leq 20. \tag{8.5}$$

Suppose that we start at

$$h(0) = 1000,\ v(0) = 0,\ m(0) = 1000. \tag{8.6}$$

A final time T is called feasible if there exists a control function $u(t)$ satisfying constraints (8.5) such that $m(t) \geq 0,\ h(t) \geq 0$ for all $t \in [0, T]$ and $h(T) = 0,\ v(T) = 0$, i.e., the lander lands softly on the moon surface. We aim at minimizing the amount of the fuel used, that is to maximize the mass once the object has landed. In other words, the objective is

$$\min_{T,u} -m(T)$$

with feasible T and u. The parameter p is unknown and needs to be estimated during operation.

Preliminary analysis. This is a problem with free final time. By using the Pontryagin Minimum Principle (PMP), we can show that the optimal control is a bang-bang control (see Example 2.3, Chapter 2). First the control exerts no thrust, letting the lander move downwards solely under gravitation. At a suitable time, maximal thrust is applied in order to brake and land the object on the surface on time. The time to begin thrusting, which depends on p, is of fundamental importance.

A serious problem occurs when we do not know the coefficient p beforehand. If the estimate of p is smaller than the true value, we would brake too early. The lander would be above the surface while the velocity approaches 0. If there is enough fuel left, we still

have a chance to smoothly land on the surface but optimality is surely lost. On the other hand, if the estimate of p is larger than the true value, we would brake too late. The lander hits the surface with a nonzero velocity. Both types of infeasibility are dangerous and may cause crashes in practice. While trying to optimize the objective function, it is of vital importance to obtain a good estimate of p on time in order to operate feasibly.

NMPC setup. Suppose that the true parameter and its initial guess together with the a priori variance are $p_{\text{true}} = 1.0,\ p_0 = 1.3,\ \sigma_0^2 = 1$, respectively. The measurements consist of the altitude and the velocity,

$$y(t) = (h(t), v(t)) + \varepsilon(t),$$

where $\varepsilon(t) \sim \mathcal{N}(0, 0.01)$ for all t. To illustrate the performance of dual NMPC, we use a penalty method to the OCP instead of dealing with the end-point constraints directly. Henceforth, we consider the following OCP

$$\min_{u,T} J(u,T) = -m(T) + 5[h^2(T) + v^2(T)]. \tag{8.7}$$

subject to system (8.4) with initial conditions (8.6) and control constraints (8.5). The function $J(u,T)$ is called *total objective*.

First we transform the problem into a problem on the fixed time interval $[0,1]$ and discretize the control $u(t)$ by piecewise constant functions. We then apply the batch NMPC strategy with $N = 40, N_c = 4$. That means for each simulation, 10 NMPC iterations are carried out. For dual control, only $N_d = 8$ future measurements are used for the OED problem.

Numerical results. By the perfect solution we mean the one obtained by solving a single OCP (8.7) subject to (8.4)–(8.5) with the true parameter k_{true}. For the perfect case, the optimal time $T_{\text{p}} = 18.95$ and the optimal mass $m_{\text{p}} = 830.54$. Moreover, $h(T_{\text{p}}) = 0.0074$, $v(T_{\text{p}}) = -0.1389$, i.e., the violation of end time constraints is tolerable. Note that because of numerical approximation, the computed control u is not truly bang-bang as the analytic one, see the first column of Figure 8.4.

The nominal NMPC solution (the first column of Figure 8.4) has a structure similar to the perfect one. The lander first uses no thrust, leading to no improvement in estimating the thrust coefficient p. Since $k_0 > k_{\text{true}}$, compared with the perfect solution, nominal NMPC applies the thrust a little late and with less magnitude. Unfortunately, this small discrepancy has a huge impact on the performance, resulting in infeasibility. The nominal NMPC solutions read as $T_{\text{n}} = 18.22$ with $h_{\text{n}}(T_{\text{n}}) = -36.50,\ v_{\text{n}}(T_{\text{n}}) = -15.88$. This likely leads to a crash of the lander. It is worth noting that once the thrust is used, the accuracy of the estimate is quickly improved (see the last row of Figure 8.4).

We now consider the performance of penalization dual NMPC (the right column of Figure 8.4), as described in Section 6.4.1. We choose α ranging from 0.1 to 2. One can observe that, the variance of the estimates of p decreases rapidly. Table 8.2 presents some statistics of numerical results. We compute means and corresponding standard deviations of 30 NMPC runs for each approach. By violation, we mean the difference of final states $(h(T), v(T))$ from the desired values $(0,0)$ in 1-norm. One can observe that dual NMPC performs better than nominal NMPC. In addition, its performance depends on the choice of the weight α which needs to be investigated carefully.

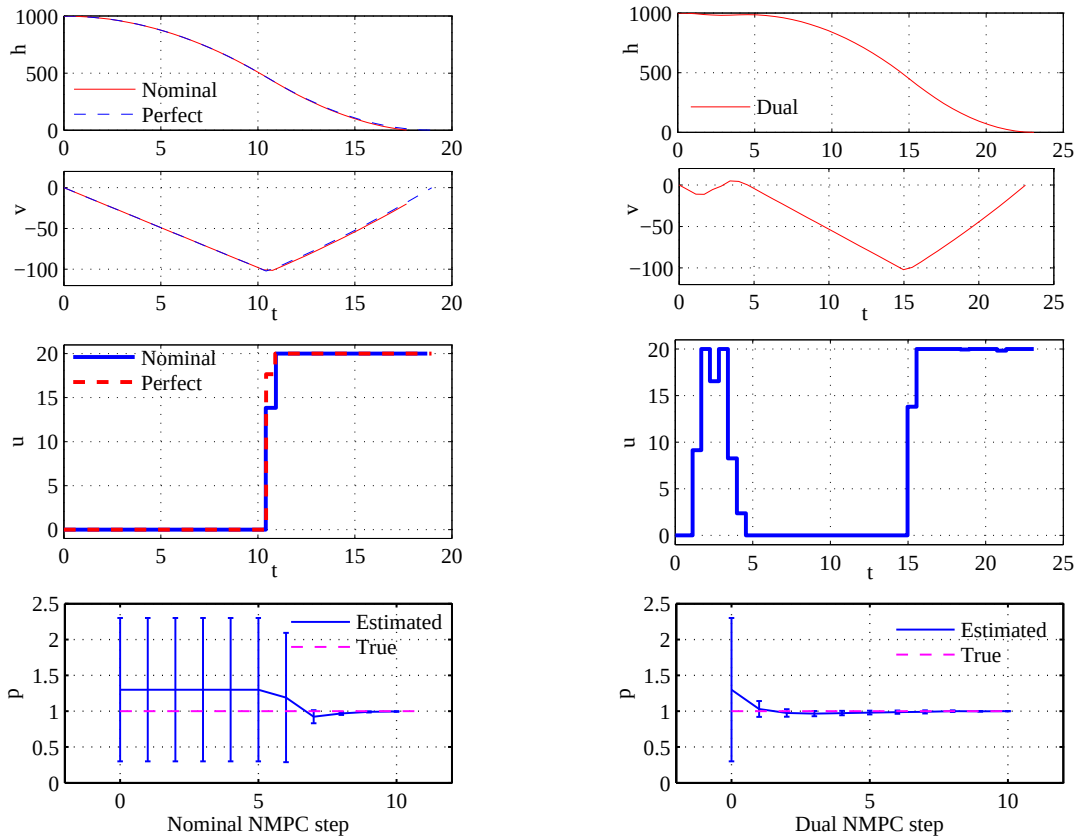

Figure 8.4: Compared to the perfect case, nominal NMPC uses thrust u a little late and with smaller magnitude, causing unacceptable infeasibility and a bad overall performance. Dual NMPC uses thrust u at suitable level at the beginning and ensures acceptable infeasibility. In the two plots at the bottom, parameter estimates are plotted with symmetric error bars of two times standard deviation. Dual NMPC is able to estimate the parameter on time while nominal NMPC is not.

Type of NMPC	**Total objective**	**m(T)**	**T**	**Violation**
Perfect	-830.46	830.55	18.96	0.14
Nominal	1368 ± 495.01	859.40 ± 2.84	17.60 ± 0.12	21.20 ± 2.11
Dual ($\alpha = 0.1$)	-570.2 ± 581.50	817.09 ± 10.56	21.02 ± 0.81	4.35 ± 5.71
Dual ($\alpha = 1$)	-682.80 ± 249.32	802.72 ± 9.09	22.70 ± 0.83	2.91 ± 4.05
Dual ($\alpha = 2$)	-766.09 ± 64.10	795.72 ± 7.51	23.46 ± 0.09	1.57 ± 1.99

Table 8.2: Statistics on the performance of various control scenarios: Dual NMPC is observed to perform better than nominal NMPC in ensuring feasibility and a good total objective. With increasing weights α, dual NMPC yields better satisfaction of the constraints but might degrade the performance.

Remark 8.3. (Interplay between performance control and information gain) *In the example above, performance control and information gain are totally conflicting at the beginning. The former keeps the thrust at zero while the latter applies the thrust maximally. Nominal NMPC exercises only performance control leading to infeasibility. In contrast to that, dual NMPC takes action to balance the two tasks suitably.*

8.3 A batch bioreactor

The model (Srinivasan et al. [81]) reads as

$$\begin{aligned}
\dot{X}(t) &= \mu(S(t))X(t) - \frac{u(t)}{V(t)}X(t),\\
\dot{S}(t) &= -\frac{\mu(S(t))X(t)}{Y_x} - \frac{vX(t)}{Y_p} + \frac{u(t)}{V(t)}(S_{\text{in}} - S(t)),\\
\dot{P}(t) &= vX(t) - \frac{u(t)}{V(t)}P(t),\\
\dot{V}(t) &= u(t)
\end{aligned} \tag{8.8}$$

where $\mu(S) = \frac{\mu_m S}{K_m + S + (S^2/K_i)}$. The states and inputs are explained in Table 8.3.

Notation	Description	Unit
X	Concentration of biomass	g/l
S	Concentration of substrate	g/l
P	Concentration of the product	g/l
V	Volume	l
S_{in}	Inlet substrate concentration	g/l
u	Flow rate of S	l/h

Table 8.3: Description of variables in a batch bioreactor

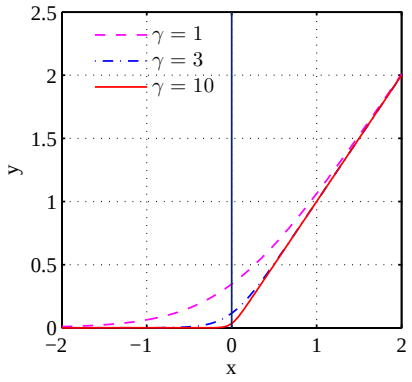

Figure 8.5: With increasing values of the smoothing parameter γ, the function $f_\gamma(x)$ better approximates the function $\max\{0, x\}$. The choice $\gamma = 10$ yields a satisfactory approximation.

The parameter Y_x is uncertain. The other constants are known with values $Y_p = 1.2, \mu_m = 0.1, K_m = 0.05, K_i = 5, v = 0.004, S_{\text{in}} = 200$. The control u is subject to constraints $0 \leq u(t) \leq 1$. The biomass $X(t)$ must not exceed 3.7. We pursue a penalization strategy to enforce this constraint. To this end, we find an approximation of the function $\max\{0, x - 3.7\}$. First we approximate the absolute value function by

$$|x|_\gamma = \frac{1}{\gamma}\log\Big[\exp(\gamma x) + \exp(-\gamma x)\Big]$$

where $\gamma > 0$ is a smoothing parameter. Then the function $f_\gamma(x) = \frac{1}{2}(x + |x|_\gamma)$ can be used to approximate the function $\max\{0, x\}$. Figure 8.5 shows that the choice $\gamma = 10$ yields a satisfactory approximation. We then augment the system with an auxiliary variable $X_a(t)$ by

$$\dot{X}_a(t) = f_\gamma(X - 3.7),$$

where $\gamma = 10$. The goal is to minimize the function J with

$$J = -P(t_f) + \beta X_a(t_f)$$

subject to (8.8) and control constraint $0 \le u(t) \le 1$ for $0 \le t \le t_f$, where $t_f = 120,\ \beta = 0.0085$. That is, we aim to maximize the product P at the end time t_f and suppress the violation of constraint $X(t) \le 3.7$ for all $t_0 \le t \le t_f$. The initial values are $X(0) = 1.0,\ S(0) = 0.5,\ P(0) = 0.0,\ V(0) = 150$.

The NMPC setting is as follows: Suppose that the true value of Y_x is 5.0. The initial guess of Y_x is $Y_x^0 = 6.0$. As a measurement function we choose $\eta = X$. The noise variance is $\sigma^2 = 0.001$. We use the sampling time $\Delta t = 2h$ and carry out 12 NMPC steps with the prediction horizon $N = 30$ and control horizon $N_c = 5$.

To assess the performance of NMPC methods, we first consider the *perfect case* when the true parameter and states are known. Our simulation, see the right column of Figure 8.6, indicates that state constraint $X(t) \le 3.7$ is nicely satisfied. The objective value is $J^*(t_f) = -1.2301$ with $P^*(t_f) = 1.2417$.

Nominal NMPC (the left column of Figure 8.6) appears to be volatile. It results in divergence of parameter and state estimates with shrinking variances (to be discussed below). Consequently, the performance of nominal NMPC is seriously worsened with $J_n(t_f) = -1.0231, P_n(t_f) = 1.0321$, i.e., 16.9% worse than the perfect case.

In contrast to that, dual NMPC (the right column of Figure 8.6) exhibits robustness, keeping the estimate close to the true values. We use penalization dual NMPC with $\alpha = 125$ in (6.19) and $N_d = 10$ in (6.16). The control exhibits excitation at the beginning. Enough information is gained to improve the estimates of the parameter. After step 5, the estimates of Y_x become reliable. This ensures good fulfillment of constraints and a good overall performance with $J_d(t_f) = -1.2094, P_d(t_f) = 1.2183$, i.e., only 1.9% worse compared to the perfect case.

In this example, MHE in nominal NMPC behaves somewhat abnormally. We have examined the estimation process carefully and discovered that, with more than 30 measurements, the parameter estimation problem as formulated by (6.21) has local minima. For example, problem (6.21) at step 5, i.e., $i = 25$, has a local minimum at $Y_x = 6.94$. If the initial guess of Y_x is slightly greater than 6.0, the computed estimate is attracted by this local minimum. Because the estimates of Y_x are wrong, the estimates of S somehow try to compensate that, see the equation of $\dot{S}(t)$ in (8.8). If we use the current estimated parameter and states and their covariance for regularization, as presented in (6.21), which are now far from the true values, MHE eventually results in divergence. This divergence behavior resembles a drawback of Kalman-filter-like approaches: Covariance matrices get unreasonably small when the estimates are far from the true values, see Grewal and Andrews [33], Chapter 8.

The analysis above leads us to a modified estimation strategy. Instead of regularization as in (6.21) for MHE, we solve least-squares problems using the full set of measurements, i.e.,

$$\min_{\hat{x}_0} \frac{1}{2} \sum_{k=0}^{i+N_c} \|y(t_k) - \eta(\hat{x}(t_k))\|^2,$$

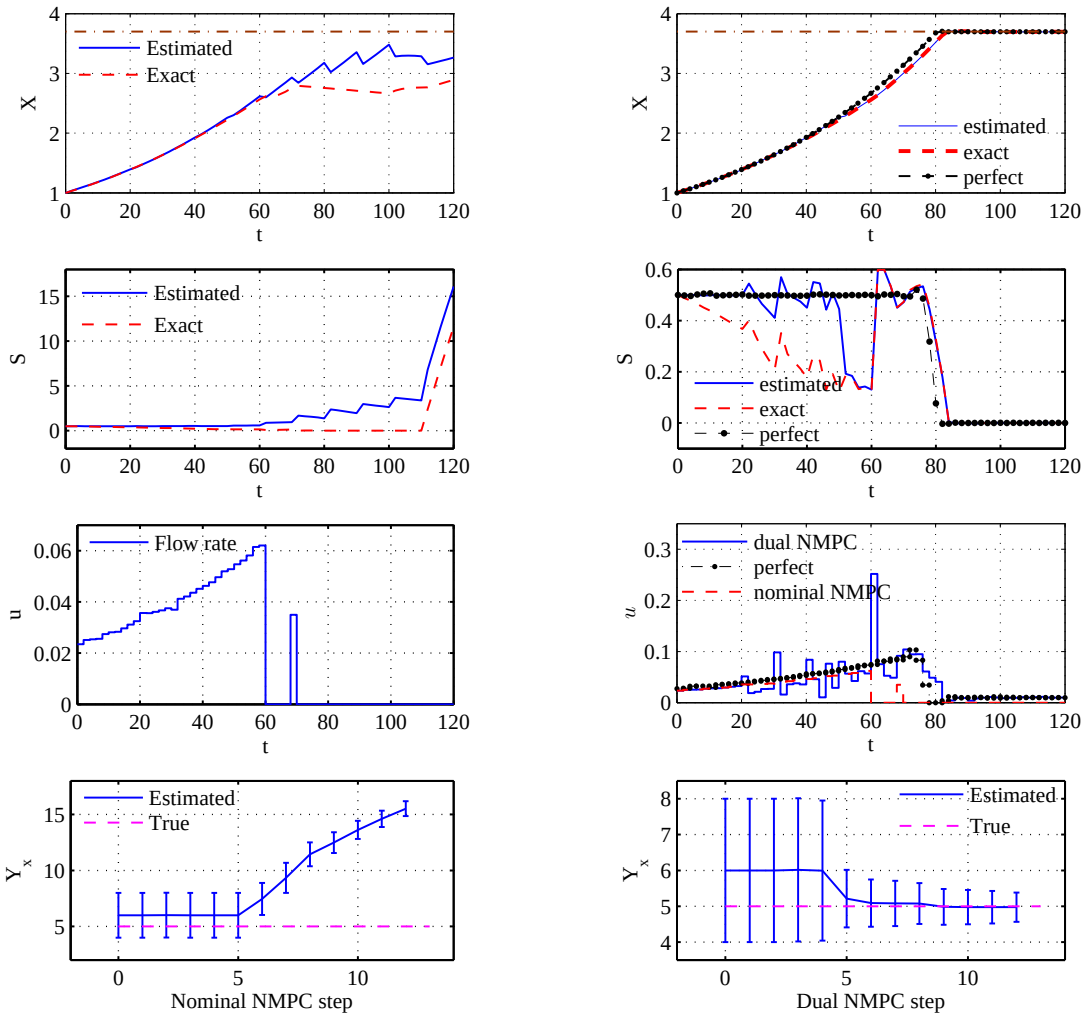

Figure 8.6: Nominal NMPC (left-hand side) leads to poor estimates of states and parameters. The objective value is significantly degraded. Dual NMPC (right-hand side) yields a control u which excites the process to estimate the parameter, ensuring constraint satisfaction and a good overall objective value. Note the excitation of dual control in the interval $t \in [0, 60]$ in comparison with nominal control. On the two plots at the bottom, parameter estimates are plotted with symmetric error bars of two times standard deviation. Dual NMPC is able to estimate the parameter on time while nominal NMPC results in divergence of parameter estimates. Here by *exact* states we mean states computed by solving the IVP with true parameters and true initial values and given control.

when $i \geq 2$. Parameter estimates, see Figure 8.7, in steps 6 and 7 move away but then converge to the true value. Despite that, nominal NMPC still noticeably underperforms dual NMPC with $J(t_f) = -1.0390$, i.e., 15.5% worse than the perfect case.

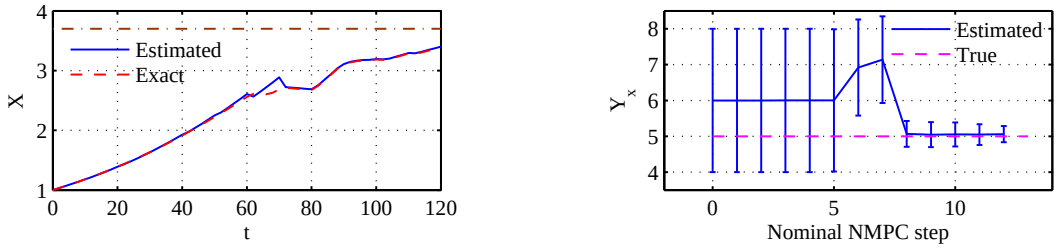

Figure 8.7: Without regularization for MHE, after moving away at some steps, parameter estimates produced by nominal NMPC converge to the true value. Nevertheless, the estimates of the parameter and states are improved so late that the performance is poor.

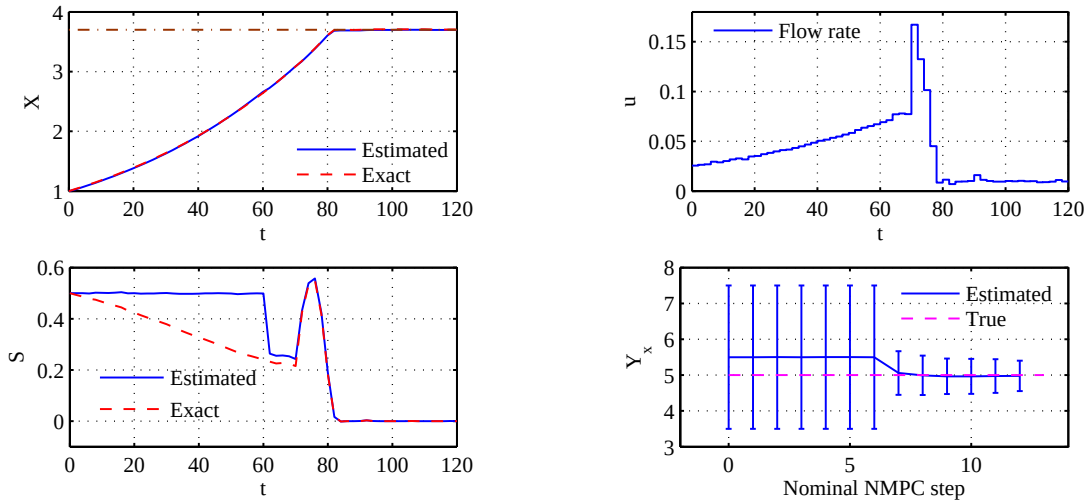

Figure 8.8: Starting with the initial guess of $Y_x^0 = 5.5$, nominal NMPC is able to estimate parameter Y_x with sufficient accuracy and yields an acceptable performance and satisfaction of the state constraint.

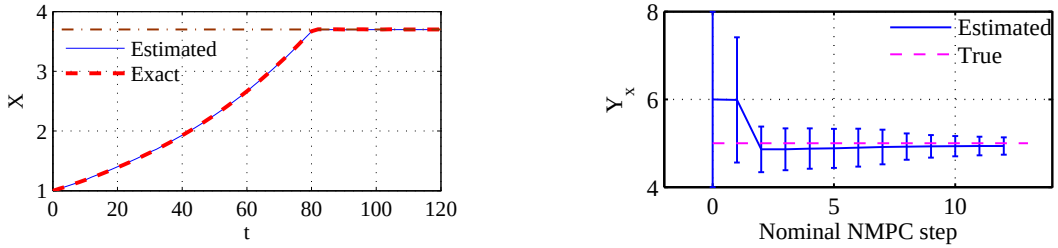

Figure 8.9: With the measurement function $\eta = (X, S)$, nominal NMPC yields good estimates for parameter and states, leading to good performance as well as constraint satisfaction. The computed objective function is $J_n(t_f) = -1.2292$, i.e, within 1.0% deviation from the perfect case

On the other hand, if we start with the estimate of $Y_x^0 = 5.5$, nominal control gives an acceptable performance, see Figure 8.8. The parameter estimates converge to the true value. It reveals that the performance of nominal NMPC may be sensitive to the initial guesses and dual NMPC can overcome this weakness.

We would like to emphasize that the measurement $\eta = X$ is not sufficiently informative, causing difficulties for nominal NMPC to estimate S and Y_x. With the choice $\eta = (X, S)$, our simulation shows that nominal NMPC performs well with precise estimates of the parameter and states, see Figure 8.9.

Remark 8.4. *In this example, we consider controls as the only degree of freedom for the OED problem. However, the OED problem for dual control can be formulated to include the choice of measurement components, see Chapter* 4 *or La et al.* [55]*. For instance, in Example* 8.3*, we could use only one sensor and choose to measure X and S at NMPC steps alternately. This leads to a more sophisticated method and is the subject of future research.*

8.4 A tractor passing a corner: Errors in measurements make controls active

We consider a realistic example of a tractor passing a corner. The model is similar to models considered in Ljung [57] and Frasch et al. [31].

$$
\begin{aligned}
\dot{x}_1 =& x_4 \cos x_3 - x_5 \sin x_3,\\
\dot{x}_2 =& x_4 \sin x_3 + x_5 \cos x_3,\\
\dot{x}_3 =& x_6,\\
\dot{x}_4 =& x_5 x_6 + \frac{1}{m}\Bigg\{ C_x(u_1+u_2)\cos u_5 - C_A x_4^2\\
&- 2C_y\left(u_5 - \frac{x_5 + d_a x_6}{x_4}\right)\sin u_5 + C_x(u_3+u_4)\Bigg\},\\
\dot{x}_5 =& -x_4 x_6 + \frac{1}{m}\Bigg\{ C_x(u_1+u_2)\sin u_5\\
&+ 2C_y\left(u_5 - \frac{x_5 + d_a x_6}{x_4}\right)\cos u_5 + 2C_y\frac{d_b x_6 - x_5}{x_4}\Bigg\},\\
\dot{x}_6 =& \frac{1}{(0.5(d_a+d_b))^2 m}\Bigg\{ d_a C_x(u_1+u_2)\sin u_5\\
&+ 2C_y\left(u_5 - \frac{x_5 + d_a x_6}{x_4}\right)\cos u_5 - 2d_b C_y\frac{d_b x_6 - x_5}{x_4}\Bigg\}.
\end{aligned}
$$

The notations are explained in the Tables 8.4–8.6. Note that COG stands for center of gravity. The tractor is controlled to follow the set-points as depicted in Figure 8.10.

Preliminary analysis. The lateral tire stiffness C_y is a key parameter. If the estimate of C_y is larger than the true value when the tractor is near the corner, the computed control actions will cause the tractor to move outwards. In case C_y is underestimated, the

State	Description	Unit
x_1	Global X-Position	m
x_2	Global Y-Position	m
x_3	Global vehicle orientation	rad
x_4	Longitudinal velocity	m/s
x_5	Lateral velocity	m/s
x_6	Yaw rate	rad/s

Table 8.4: Description of the states in the vehicle example

Parameter	Description	Value	Unit
m	Vehicle mass	1500	kg
C_A	Air resistance	0.5	m^{-1}
d_a	front axle to COG	1.5	m
d_b	rear axle to COG	1.5	m
C_x	Longitudinal tire stiffness	Unknown	N
C_y	Lateral tire stiffness	Unknown	N/rad

Table 8.5: Description of the parameters in the vehicle example

Input	Description	Value	Unit
u_1	Slip of front left	To be computed	–
u_2	Slip of front right	To be computed	–
u_3	Slip of rear left	0	–
u_4	Slip of rear right	0	–
u_5	Steering angle	To be computed	rad

Table 8.6: Description of the inputs in the vehicle example

tractor will move deeply inwards. With this reasoning we deduce that it is advantageous to have a good estimate of the lateral tire stiffness before entering the corner.

Moreover the steering angle directly affects the identifiability of the lateral tire stiffness. The information gain task should make the tractor wiggle around the desired path. But this is opposed to the performance control task which aims to track the path. Therefore it is decisive to strike a balance between the two tasks.

NMPC setup. We suppose that u_1 and u_2 are the same and call them both by slip of front tire. Thus there are two controls u_1, u_5. Additionally, we consider constraints on them given by

$$0 \leq u_1 \leq 0.1, \quad -1 \leq u_5 \leq 1.$$

Assume that the true parameters are $C_x^* = 200, C_y^* = 25$. For the measurement function we consider

$$\eta(x) = (x_1, x_2)^T$$

and white noise with variance 0.01.

The tractor starts at $(X_0, Y_0) = (-20, 0)$ with initial conditions $(-20, 0, 0, 10, 0, 0)$. We carry out 6 NMPC steps with a prediction horizon of $N = 10$ and a control horizon

of 3. The covariance matrix used in dual control is computed based on $N_d = 6$ future measurements. The desired path is constructed by $Z = (X^*, Y^*)$ which represents the set-points in Figure 8.10: $X_i^* = -20 + 2i, i = 0, 1, ..., 8$, $X_9^* = 2$ and $X_i^* = 0, i \geq 10$; $Y_i^* = 0, i = 0, 1, ..., 8, Y_9^* = 2$ and $Y_i^* = 4 + 2(i - 10), i \geq 10$. The objective function for the NMPC step t is

$$J = \frac{1}{2} \sum_{k=1}^{N} (X_{t+k} - X_{t+k}^*)^2 + (Y_{t+k} - Y_{t+k}^*)^2.$$

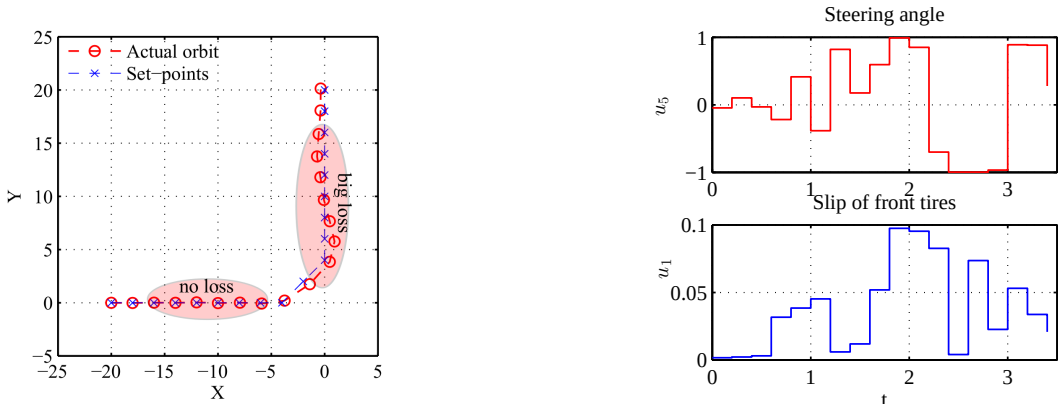

Figure 8.10: To keep the tractor on the horizontal line at the beginning, nominal NMPC exercises almost no steering. Little information is then gained to estimate the lateral tire stiffness. The tractor makes a zigzag turn at the corner.

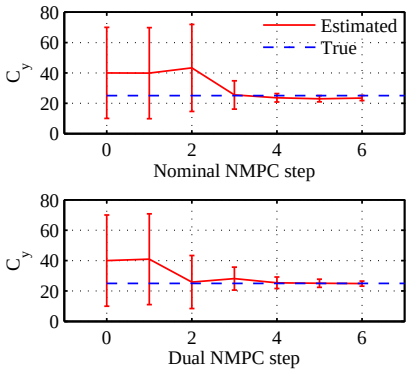

Figure 8.11: Parameter estimates are plotted with error bars which are symmetric and 2 times standard deviation long. The estimates of C_y in nominal NMPC (the upper plot) are not improved at the beginning while they quickly become accurate in dual NMPC (the lower plot).

Numerical results. Nominal NMPC (Figures 8.10–8.11), instructs the tractor to tightly follow the path at the beginning. It exercises too little steering, causing lack of information for estimating the lateral tire stiffness C_y. As a result, the estimate of C_y is inaccurate by the time the tractor enters the corner. The tractor tends to move out of

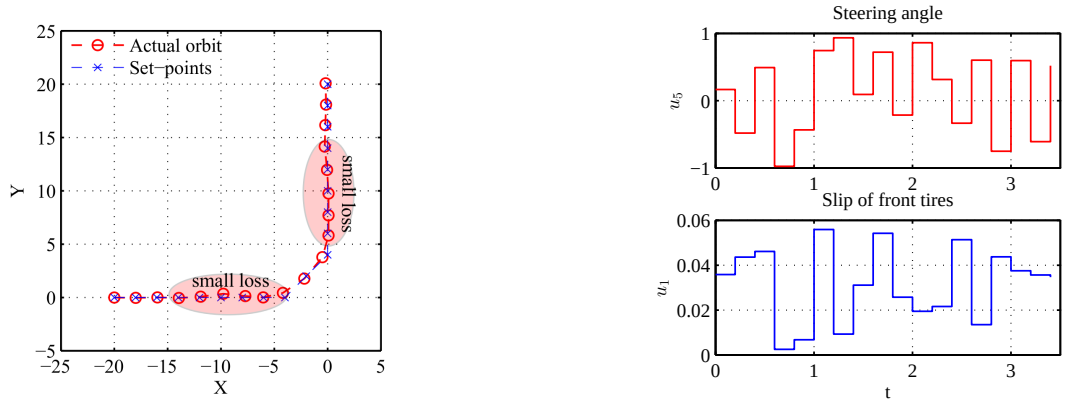

Figure 8.12: Dual NMPC creates a wiggle movement at the beginning, helping to estimate the lateral tire stiffness on time. The tractor makes a smooth turn through the corner.

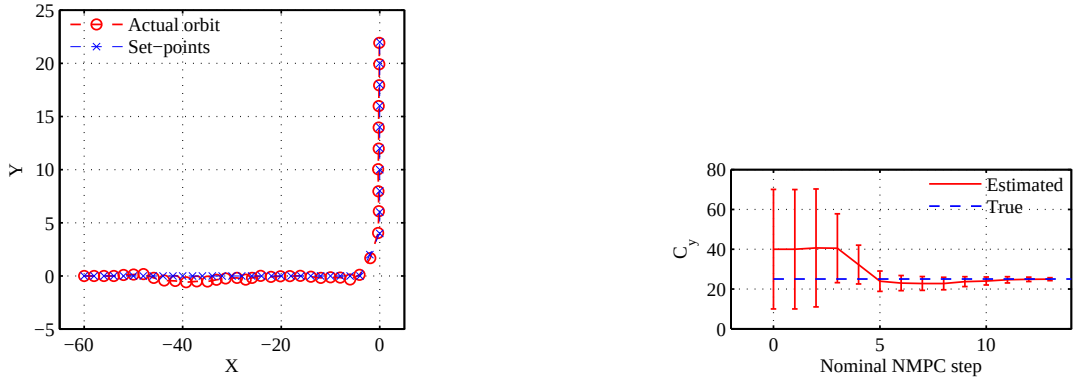

Figure 8.13: Nominal NMPC yields a satisfactory performance when starting far from the corner. This is because the measurement errors make the control active, forcing the tractor to slightly deviate from the desired path. The estimates of the parameter are gradually improved and get accurate on time.

the path wildly. After that, the estimates of C_y quickly become better and the tractor returns to the desired path quite smoothly. At the same time, there is no problem with estimating the longitudinal tire stiffness. Its estimates attain an acceptable accuracy after only 2 NMPC steps.

Dual NMPC (Figures 8.11–8.12) is observed to do a better job. It first makes a little steering around the track, although rather small but this action proves to be valuable because it helps to quickly reduce the uncertainty on the estimate of the lateral tire stiffness. The performance index, i.e., deviation from the path of dual NMPC is 1.20, and the nominal NMPC is 1.53, on average of 60 runs.

Consider now the case when we start far from the corner. At each NMPC step, we estimate the states, including x_1, x_2. The estimates are often not precise. So even if the true (x_1, x_2) does lie in the path, their estimates do not. The controller will try to steer the tractor towards the path. This introduces the wiggling effect and contributes to the estimation of the lateral tire stiffness. Our numerical results (Figures 8.13) indicate that in case we start at $(X_0, Y_0) = (-60, 0)$, the lateral tire stiffness C_y is identified on

time, resulting in a good performance. On the other hand, if there were perfect state estimates and the tractor was on the horizontal part of the path, then there would be no improvement on the estimates of C_y. So interestingly, measurement noise acts as an excitation helping to increase the accuracy of parameter estimates.

All computations are done in MATLAB with the MEX Interface of SolvIND (Albersmeyer [1], Albersmeyer and Bock [3]), a software package developed at the group Simulation and Optimization, IWR, University of Heidelberg. SolvIND provides ODE solvers and sensitivities of solutions with respect to parameters and initial conditions using ADOL-C (Walther and Griewank [85]). The derivatives computed by SolvIND are given to `fmicon` and `lsqnonlin` from MATLAB to solve OCPs and estimation problems. Random noise in measurements is generated by the function `randn`.

Conclusions and Outlook

In this work, we developed dual control methods for controlling uncertain processes. The core idea is to strike a balance between the goal of optimizing the cost function and the goal of estimating uncertain parameters. Using Optimal Experimental Design (OED), we proposed novel approaches in the framework of Nonlinear Model Predictive Control (NMPC). Our methods are superior to conventional control methods in ensuring feasibility, optimizing economic objectives and enhancing robustness. This is illustrated by a collection of examples from vehicle control and chemical engineering. We also investigated the asymptotic behavior of the least squares methods for parameter estimation and suggested a sequential least squares strategy to deal with possibly ill-posed estimation problems. In addition, we made contributions to the stability analysis of NMPC by extending partial stability results of ordinary differential equations for NMPC. As continuation of the dissertation research, we would like to investigate the following issues.

Optimal choice of weight α

Through the numerical examples in Chapter 8, we have seen the effect of weight α on the performance of Dual Control. Small values of α prioritize the performance control task and lead to a similar result as nominal control. Big values of α overemphasize the information gain task and may worsen the overall performance. It is therefore important to choose the weight α in an optimal way.

In Chapter 6, we interpreted the penalty term as an approximation of the variance of the nominal objective value. One can choose α based on confidence intervals. For example, if the nominal objective value is well approximated by a normal distribution, then for a confidence level of 95%, one can choose $\alpha = 1.96$. However, those are still approximations because in most cases the nominal objective value has a complicated distribution. Hence the issue of choosing a suitable α in the general case needs to be investigated. Last but not least, depending on the magnitude of the control objective and the OED objective, α can be chosen adaptively along NMPC steps.

Dual Control for stochastic models

Our vision is to study dual control methods for stochastic models, which are becoming more and more popular. In this framework, noise and disturbances can be presented in a rigorous manner. Using tools from stochastic optimization and stochastic control, we can lay a theoretical foundation and develop more powerful and more efficient dual control methods.

Dual Control in the framework of scenario trees

The scenario tree approach has an advantage of enhancing robustness and could ensure feasibility with high probability. It is promising for controlling uncertain processes and has received intensive research in recent years, see for example, Engell [23], Heitsch and Römisch [38].

The scenario tree approach makes the assumption that the uncertain parameters assume a finite number of values with some probability distribution. Essentially, it is a stochastic approach in which the distributions of uncertainties are treated in discrete form with a small number of realizations. The branches of the scenario tree correspond to the evolution of uncertainties. This results in an exponential growth in the number of uncertain values and horizon control. As a consequence, we encounter large nonlinear optimization problems that need to be handled efficiently. If the range of uncertainty is small, the number of scenarios considered can be reduced, leading to saving in the computational efforts. As we have seen, Dual Control aims to get informative data for reliable estimates, i.e., estimates with small confidence regions. In this connection, the scenario tree approach could be combined with Dual Control to yield more efficient control methods.

Application areas

Our ultimate goal is to develop and apply dual control methods to real-life problems, particularly in biology and medical treatment. This would be challenging but is also promising. For instance, in medical treatments, there are stringent guidelines on cost efficiency, safety requirements and effectiveness of treatments. In this regard, Dual Control is believed to offer efficient solutions in making a suitable balance among those factors.

Appendix A

In this Appendix, we present results on *envelope theorems* and *transversality conditions*. They are useful for understanding the Pontryagin Minimum Principle and the Hamilton-Jacobi-Bellman Equations introduced in Chapter 2.

A.1 An envelope theorem

Let ϕ be a scalar function

$$\phi : U \times P \to \mathbb{R},$$

where U is a set, not necessary equipped with any topological or algebraic structure, and $P \in \mathbb{R}^{n_p}$ is the set of parameters. For each $p \in P$, define

$$V(p) = \inf_{u \in U} \phi(u, p), \quad U^*(p) = \{u \in U : \phi(u, p) = V(p)\}.$$

We assume that solutions exist. However, we do not assume the uniqueness of solutions. Our goal is to characterize the derivative of $V(p)$.

Theorem A.1. (Milgrom and Segal [62]) *Assume that for each $u \in U$, there exists the partial derivative $\frac{\partial \phi}{\partial p}(x, p)$. Then for any $p \in P$, $u_p \in U^*(p)$ and $h \in \mathbb{R}^{n_p}$ such that $p + th \in P$ for all $t \in \mathbb{R}$ sufficiently small, the following inequalities hold*

$$\lim_{t \to 0^+} \frac{V(p + th) - V(p)}{t} \leq \frac{\partial \phi}{\partial p}(u_p, p)h \leq \lim_{t \to 0^-} \frac{V(p + th) - V(p)}{t}.$$

If $V(p)$ is differentiable at p then

$$\frac{\partial V}{\partial p}(p) = \frac{\partial \phi}{\partial p}(u_p, p).$$

Proof. From the definition of V and U_p we have

$$V(p) = \phi(u_p, p), \quad V(p + th) \leq \phi(u_p, p + th).$$

It follows that

$$V(p + th) - V(p) \leq \phi(u_p, p + th) - \phi(u_p, p).$$

As a result, for $t_1 > 0$ and $t_2 < 0$ small enough,

$$\frac{V(p + t_1 h) - V(p)}{t_1} \leq \frac{\phi(u_p, p + t_1 h) - \phi(u_p, p)}{t_1},$$

and

$$\frac{V(p + t_2 h) - V(p)}{t_2} \geq \frac{\phi(u_p, p + t_2 h) - \phi(u_p, p)}{t_2}.$$

Letting $t_1 \to 0^+$ and $t_2 \to 0^-$, we obtain

$$\lim_{t\to 0^+} \frac{V(p+th)-V(p)}{t} \le \frac{\partial \phi}{\partial p}(u_p, p)h \le \lim_{t\to 0^-} \frac{V(p+th)-V(p)}{t}. \tag{A.9}$$

If $V(p)$ is differentiable at p then

$$\lim_{t\to 0^+} \frac{V(p+th)-V(p)}{t} = \lim_{t\to 0^-} \frac{V(p+th)-V(p)}{t} = \frac{\partial V}{\partial p}(p)h.$$

It follows from (A.9) that

$$\frac{\partial V}{\partial p}(p)h = \frac{\partial \phi}{\partial p}(u_p, p)h.$$

This holds for any h. Therefore $\dfrac{\partial V}{\partial p}(p) = \dfrac{\partial \phi}{\partial p}(u_p, p)$. The proof is complete. □

A.2 A lemma on transversality conditions

The following lemma is fundamental for the derivation of the Lagrange multipliers. In particular, it is applied to establish the boundary conditions for the adjoint variables in PMP. To state and prove the lemma, we need some preliminary knowledge from differential geometry, for example differential manifold, tangent spaces.

Suppose that $m \le n$ and $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a differentiable function. Set

$$M = \big\{x \in \mathbb{R}^n \;\big|\; h(x) = 0\big\}.$$

For $x \in M$, the *tangent space* of M at x is defined by

$$\partial_x M = \Big\{d \in \mathbb{R}^n \mid \text{ there exist } \varepsilon > 0 \text{ and a differentiable function } \varphi : (-\varepsilon, \varepsilon) \to M, \quad \varphi(0) = x, \quad \varphi'(0) = d\Big\}.$$

It is easy to see that

$$\partial_x M \subseteq \{d \in \mathbb{R}^n \mid \frac{\partial h}{\partial x}(x)d = 0\}.$$

For each regular point x of M, i.e., $\dfrac{\partial h}{\partial x}(x)$ has full rank m, the above relation holds with equality sign due to the implicit function theorem, i.e.,

$$\partial_x M = \{d \in \mathbb{R}^n \mid \frac{\partial h}{\partial x}(x)d = 0\}. \tag{A.10}$$

Now let $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ be continuously differentiable. Clearly if $f(x) = g(x)$ in an open neighborhood of x_0 then

$$\frac{\partial f}{\partial x}(x_0) = \frac{\partial g}{\partial x}(x_0).$$

We would like to investigate a similar assertion for the case when $f(x) = g(x)$ only holds in M.

Lemma A.1. (Zeidler [88]) *If $f(x) = g(x)$ for all $x \in M$, then for x_0 which is a regular point of M, there exists $\nu \in \mathbb{R}^m$, depending on x_0 such that*

$$\frac{\partial f}{\partial x}(x_0) = \frac{\partial g}{\partial x}(x_0) + \frac{\partial h}{\partial x}(x_0)\nu.$$

Proof. For any $d \in \partial_{x_0} M$, there exist $\varepsilon > 0$ and $\varphi(t) : (-\varepsilon, \varepsilon) \longrightarrow M$ such that

$$\varphi(0) = x_0, \quad \varphi'(0) = d.$$

Since $\varphi(t) \in M$, by the hypothesis

$$f(\varphi(t)) - g(\varphi(t)) = 0, \quad t \in (-\varepsilon, \varepsilon).$$

Differentiating both side with respect to t at x_0 yields

$$\left(\frac{\partial f}{\partial x}(x_0) - \frac{\partial g}{\partial x}(x_0) \right) d = 0.$$

Hence

$$\left(\frac{\partial f}{\partial x}(x_0) - \frac{\partial g}{\partial x}(x_0) \right) \in (\partial_{x_0} M)^{\perp},$$

where $(\partial_{x_0} M)^{\perp}$ denotes the orthogonal complement of $\partial_{x_0} M$ in $\mathbb{R}^n$. Since x_0 is a regular point of M, from (A.10) we have

$$(\partial_{x_0} M)^{\perp} = \text{ span} \left\langle \frac{\partial h}{\partial x}(x_0) \right\rangle.$$

As a consequence, there exists $\nu \in \mathbb{R}^m$ such that

$$\frac{\partial f}{\partial x}(x_0) - \frac{\partial g}{\partial x}(x_0) = \frac{\partial h}{\partial x}(x_0)\nu.$$

This completes the proof. □

Appendix B

This Appendix provides some basic facts about the problem of parameter and state estimation for the linear case. They include the Gauss-Markov theorem, conditional expectation as the optimal estimator with respect to the least-squares criterion, the equivalence of linear and nonlinear estimates in the Gaussian case. Our presentation also tries to connect the estimation problem with some closely related problems from the inverse and ill-posed realms, see Chapter 4, Vogel [83].

B.1 Best linear unbiased estimator and the Gauss-Markov theorem

The model under consideration reads as

$$y = Ax + \varepsilon, \tag{B.11}$$

where $A \in \mathbb{R}^{m\times n}$ is a deterministic matrix, $\varepsilon \in \mathbb{R}^m$ are random noise, $y \in \mathbb{R}^m$ are noisy measurements; $x \in \mathbb{R}^n$ is the parameter to be estimated which is deterministic. Suppose that ε has mean zero and covariance $R \succ 0$. No assumption on the distribution type of ε is required. Our task is to estimate x from the measurements y. In order to assess how good an estimate is, we specify some criteria on the error estimate $e = \hat{x} - x$. For this purpose, consider functions $L : \mathbb{R}^n \to \mathbb{R}^+$, such as,

$$L(z) = \|z\|^2, \quad L(z) = |z|_1, \quad L(z) = \exp(\|z\|).$$

In this section we confine ourselves to the mean squares error criterion,

$$\hat{x} = \min_Z \mathbb{E}\,\|Z - x\|^2 .$$

Definition B.1. (Vogel [83]) *An estimate $\hat{x}$ of x is called unbiased if $\mathbb{E}(\hat{x}) = x$.*

By an estimator, we understand a Borel function $f : \mathbb{R}^m \to \mathbb{R}^n$, which gives for given data y a unique estimate $f(y)$ of x.

We assume that $m \geq n$ and A has full rank, i.e., $\operatorname{rank}(A) = n$. This implies that the matrix $A^T DA$ is invertible for and positive definite matrix $D \in \mathbb{R}^{m\times m}$. In case the estimator is linear in y, i.e., f has the form of a matrix $B \in \mathbb{R}^{n\times m}$, we have the explicit solution, given by the Gauss-Markov theorem.

Theorem B.2. (Vogel [83]) *The best linear unbiased estimator (BLUE) is given by*

$$\hat{x} = \hat{B}y, \quad \hat{B} = (A^T R^{-1} A)^{-1} A^T R^{-1}.$$

Moreover, for any linear unbiased estimate z of x we have

$$\operatorname{Cov}(z) \succeq \operatorname{Cov}(\hat{x}).$$

Proof. Suppose that $B \in \mathbb{R}^{n\times m}$ represents an unbiased estimator. Because of unbiasedness,

$$\mathbb{E}(BY) = \mathbb{E}(BAx + \varepsilon - x) = (BA - \mathbb{I}_n)x = 0$$

since $\mathbb{E}\varepsilon = 0$. That holds for any $x \in \mathbb{R}^n$, hence $BA = \mathbb{I}_n$. Obviously, $\hat{B}A = \mathbb{I}_n$. Set $H = B - \hat{B}$, then $HA = 0$. We have

$$\begin{aligned}\mathbb{E}\,\|By - x\|^2 =& \mathbb{E}\left\|(\hat{B} + H)(Ax + \varepsilon) - x\right\|^2 = \mathbb{E}\left\|(\hat{B} + H)\varepsilon\right\|^2\\ =& \mathbb{E}\varepsilon^T(\hat{B} + H)^T(\hat{B} + H)\varepsilon\\ =& \mathbb{E}\varepsilon^T\hat{B}^T\hat{B}\varepsilon + \mathbb{E}\varepsilon^T H^T H\varepsilon + 2\mathbb{E}\varepsilon^T H^T\hat{B}\varepsilon\\ =& \mathbb{E}\left\|\hat{B}y - x\right\|^2 + \mathbb{E}\varepsilon^T H^T H\varepsilon + 2\mathbb{E}\varepsilon^T H^T\hat{B}\varepsilon.\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E}\varepsilon^T H^T\hat{B}\varepsilon =& \mathbb{E}(\operatorname{Trace} H^T\hat{B}\varepsilon\varepsilon^T) = \operatorname{Trace} H^T\hat{B}\mathbb{E}(\varepsilon\varepsilon^T) = \operatorname{Trace}(H^T\hat{B}R)\\ =& \operatorname{Trace}(H^T(A^TR^{-1}A)^{-1}A^T) = \operatorname{Trace}(A^TR^{-1}A)^{-1}A^TH^T = 0\end{aligned}$$

since $HA = 0$. Therefore

$$\mathbb{E}\,\|By - x\|^2 = \mathbb{E}\left\|\hat{B}y - x\right\|^2 + \mathbb{E}\varepsilon^T H^T H\varepsilon \geq \mathbb{E}\left\|\hat{B}y - x\right\|^2.$$

This proves the optimality of $\hat{B}$. Moreover, the equality holds if and only if $\mathbb{E}\varepsilon^T H^T H\varepsilon = 0$. It follows that $\operatorname{Trace} H^T HR = 0$. Since $H^T H \succeq 0$ and $R \succ 0$, one can easily deduce that $H = 0$. The second part of the theorem can be proved similarly. In fact, with $z = By$, we have

$$\begin{aligned}\operatorname{Cov} z =& \mathbb{E}(By - x)(By - x)^T = \mathbb{E}(\hat{B} + H)\varepsilon\varepsilon^T(\hat{B} + H)^T\\ =& (\hat{B} + H)R(\hat{B} + H)^T = \hat{B}R\hat{B}^T + HR\hat{B}^T + \hat{B}RH^T + HRH^T\\ =& \operatorname{Cov}(\hat{B}y) + HR\hat{B}^T + \hat{B}RH^T + HRH^T.\end{aligned}$$

Again $\hat{B}RH^T = 0$ and $HR\hat{B}^T = 0$. Thus

$$\operatorname{Cov} z = \operatorname{Cov}(\hat{B}y) + HRH^T \succeq \operatorname{Cov}(\hat{B}y).$$

This completes the proof. □

Remark B.5. *The BLUE coincides with the weighted LSQ solution which solves*

$$\min_z \frac{1}{2}(y - Az)^T R^{-1}(y - Az).$$

B.2 Linear Minimum Variance Estimator and Tikhonov regularization method

If we consider x in (B.11) as random with some statistical properties, we then deal with a fully stochastic problem. It is motivated from the fact that x are states of a dynamics system which are uncertain and some a priori information about x is available. After obtaining measurements y, we want to estimate x. The view-point that x is random also

has its origin in the theory of ill-posed problems. The problem with BLUE is that the matrix $A^T R^{-1} A$ may be of large condition, leading to numerical instability. The a priori information in x helps to reduce its condition number. To emphasize that x is random, we change it to upper case and rewrite (B.11) as follows

$$Y = AX + \varepsilon. \tag{B.12}$$

The linear minimum variance estimator of X is defined by $\hat{X} = \hat{B}Y$ where $\hat{B}$ solves the following problem

$$\min_{B \in \mathbb{R}^{n \times m}} \mathbb{E}(\|BY - X\|^2).$$

Theorem B.3. (Vogel [83]) *Suppose X, Y are jointly distributed and*

$$F_{XX} = \mathbb{E}(XX^T), \quad F_{XY} = \mathbb{E}(XY^T), \quad F_{YY} = \mathbb{E}(YY^T).$$

Then $\hat{B} = F_{XY} F_{YY}^{-1}$.

Proof. The proof can be carried out straightforwardly. In fact, for any matrix $B \in \mathbb{R}^{n \times m}$, set $H = B - \hat{B}$. We have

$$\begin{aligned} \mathbb{E}\,\|BY - X\|^2 &= \mathbb{E}\left\|\hat{B}Y - X + HY\right\|^2 \\ &= \mathbb{E}\left\|\hat{B}Y - X\right\|^2 + \mathbb{E}\,\|HY\|^2 + \mathbb{E}(HY)^T(\hat{B}Y - X). \end{aligned}$$

On the other hand,

$$\begin{aligned} \mathbb{E}(HY)^T(\hat{B}Y - X) &= \mathbb{E}(Y^T H^T \hat{B}Y) - \mathbb{E}(Y^T H^T X) \\ &= \mathbb{E}\,\mathrm{Trace}(H^T \hat{B} Y Y^T) - \mathbb{E}\,\mathrm{Trace}(H^T X Y^T) \\ &= \mathrm{Trace}(H^T \hat{B} \mathbb{E} Y Y^T) - \mathrm{Trace}(H^T \mathbb{E} X Y^T) \\ &= \mathrm{Trace}(H^T \hat{B} F_{YY}) - \mathrm{Trace}(H^T F_{XY}) = 0 \end{aligned}$$

because $\hat{B}F_{YY} = F_{XY}$ by the choice of $\hat{B}$. It follows that

$$\mathbb{E}\,\|BY - X\|^2 = \mathbb{E}\left\|\hat{B}Y - X\right\|^2 + \mathbb{E}\,\|HY\|^2 \geq \mathbb{E}\left\|\hat{B}Y - X\right\|^2.$$

This completes the proof. □

Note that we define $F_{ZV} = \mathbb{E}(ZV^T)$ for random vectors Z and V instead of $\mathbb{E}(Z - \mathbb{E}Z)(V - \mathbb{E}V)^T$. In practice, the expected value of X is hardly known at the beginning. Therefore, it is difficult to determine whether the estimate is biased or not. One possibility is to use some initial guess, namely x_0 and consider it as the expected value of X. If so, we can point out that the solution corresponds to the Kalman filter in case of Gaussian distribution. In fact, suppose that

$$\tilde{Y} = A\tilde{X} + \varepsilon,$$

where $\tilde{X} \sim \mathcal{N}(x_0, P_0)$, $\varepsilon \sim \mathcal{N}(0, R)$ and $\tilde{X}$ and ε are independent. After setting

$$X = \tilde{X} - x_0, \; Y = \tilde{Y} - Ax_0$$

we have

$$F_{XX} = P_0,\ F_{XY} = P_0 A^T,\ F_{YY} = AP_0A^T + R.$$

The formula for the optimal estimate is

$$\hat{X} = P_0A^T(AP_0A^T + R)^{-1}Y,$$

or in terms of the original variables

$$\hat{\tilde{X}} = x_0 + P_0A^T(AP_0A^T + R)^{-1}(\tilde{Y} - Ax_0).$$

which agrees with the Kalman update.

B.3 Conditional distribution and optimality of the linear estimator

We consider estimators which are Borel functions of data, $f : \mathbb{R}^m \to \mathbb{R}^n$, $z = f(y)$. In case of linear estimators, f corresponds to some matrix $B \in \mathbb{R}^{n\times m}$. Because the solution of the linear case is computationally favorable, we could naturally ask, whether we can get better estimate by enlarging the set of admissible estimates. We will see later that in case of Gaussian distributions, the optimal nonlinear estimate coincides with the linear solution. However, in general this does not hold. We first characterize the optimal estimators in case of general distributions, then will give their explicit form for the Gaussian case.

Denote the class of Borel functions from $\mathbb{R}^m$ into $\mathbb{R}^n$ by $\mathcal{B}$. The estimation problem can be stated as

$$\min_{f\in\mathcal{B}} \mathbb{E}\big(\,\|f(Y) - X\|^2\,\big). \tag{B.13}$$

Lemma B.2. *Suppose $\hat{f}$ is a solution of* (B.13)*. Then for any $g \in \mathcal{B}$, it holds that*

$$\mathbb{E}\big[(\hat{f}(Y) - X)^T g(Y)\big] = 0.$$

Proof. Consider the function $\varphi :\ \mathbb{R} \to \mathbb{R}$ defined by

$$\varphi(t) = \frac{1}{2}\mathbb{E}\left\|(\hat{f} + tg)(Y) - X\right\|^2.$$

From the hypothesis, $\varphi(t)$ assumes a minimum at $t = 0$. Hence $\frac{d\varphi}{dt}(0) = 0$. On the other hand

$$\varphi(t) = \frac{1}{2}\mathbb{E}\left\|\hat{f}(Y) - X\right\|^2 + \frac{1}{2}t^2\mathbb{E}\,\|g(Y)\|^2 + t\mathbb{E}\left[(\hat{f}(Y) - X)^T g(Y)\right].$$

It follows that

$$\frac{d\varphi}{dt} = t\mathbb{E}\,\|g(Y)\|^2 + \mathbb{E}\left[(\hat{f}(Y) - X)^T g(Y)\right].$$

Consequently

$$\mathbb{E}\left[(\hat{f}(Y) - X)^T g(Y)\right] = \frac{d\varphi}{dt}(0) = 0.$$

The proof is complete. □

Due to Lemma (B.2), if two random vectors $\hat{f}(Y)$ and $\hat{g}(Y)$ both solve (B.13) then $\hat{f}(Y) = \hat{g}(Y)$ almost everywhere. This also implies that $\hat{f} = \hat{g}$ almost everywhere in $\mathbb{R}^m$.

The conditional expectation of X given Y can be defined as a random vector $g(Y)$ where $g : \mathbb{R}^m \to \mathbb{R}^n$ is a Borel function which solves (B.13), denoted by $\mathbb{E}\big[X|Y\big]$, see Durrett [22]. With this setup, we have claimed that the solution of the estimation problem is the conditional expectation of X given Y.

If there is no restriction on f, it is impossible to determine the optimal solution of (B.13). However for the Gaussian case, we can obtain the optimal estimator explicitly. It turns out that in case of Gaussian distributions, the optimal linear estimator is also optimal in the set of all linear and nonlinear estimators. To show this, we need the following lemma. Recall that the covariance matrices of random vectors Z and V by

$$\operatorname{Cov} Z = \mathbb{E}(Z - \mathbb{E}Z)(Z - \mathbb{E}Z)^T, \quad \operatorname{Cov}(Z, V) = \mathbb{E}(Z - \mathbb{E}Z)(V - \mathbb{E}V)^T.$$

Lemma B.3. (Kalman [44]) *Suppose that Z and Y are random vectors with values in $\mathbb{R}^m$, $\mathbb{R}^n$, respectively, which are jointly Gaussian and uncorrelated, i.e.,*

$$\operatorname{Cov}(Z, Y) = 0,$$

then Z and Y are independent.

By independence of random vectors, we mean that their joint density function is equal to the product of their marginal density functions, $f_{(Z,Y)}(z, y) = f_Z(z) f_Y(y)$.

Proof. Without loss of generality, assume that $\mathbb{E}Z = 0, \mathbb{E}Y = 0$. Then $\mathbb{E}(ZY^T) = \mathbb{E}(YZ^T) = 0$. Since (Z, Y) is jointly Gaussian distributed, its density function has the form

$$f_{ZY}(z, y) = \frac{1}{\sqrt{(2\pi)^{m+n} \det \Sigma_{ZY}}} \exp\left\{-\frac{1}{2}(z \quad y)^T \Sigma_{ZY}^{-1} \begin{pmatrix} z \\ y \end{pmatrix}\right\}.$$

where

$$\Sigma_{ZY} = \begin{pmatrix} \Sigma_Z & 0 \\ 0 & \Sigma_Y \end{pmatrix}, \quad \Sigma_Z = \operatorname{Cov} Z, \quad \Sigma_Y = \operatorname{Cov} Y.$$

We can rewrite $f_{ZY}(z, y)$ so that z and y are separated. In fact

$$\begin{aligned} f_{ZY}(z, y) = &\frac{1}{\sqrt{(2\pi)^n \det \Sigma_Z}} \exp\left\{-\frac{1}{2} z^T \Sigma_Z^{-1} z\right\} \\ &\times \frac{1}{\sqrt{(2\pi)^m \det \Sigma_Y}} \exp\left\{-\frac{1}{2} y^T \Sigma_Y^{-1} y\right\} \end{aligned}$$

It follows straightforwardly that Z and Y are normally distributed with the density functions $f_Z(z)$ and $f_Y(y)$ satisfying $f_{ZY}(z, y) = f_Z(z) f_Y(y)$. Consequently, Z and Y are independent. □

We are now in the position to show that under the assumption of a Gaussian distribution, the optimal linear estimator provides the optimal solution to (B.13).

Lemma B.4. (Kalman [44]) *Suppose that in* (B.13) *X and Y are jointly Gaussian and $\operatorname{Cov} Y$ is nonsingular. Set $B = \operatorname{Cov}(X, Y)(\operatorname{Cov} Y)^{-1}$ and $Z = BY - X$. Then $Z - \mathbb{E}Z$ is an optimal solution of* (B.13).

Proof. Obviously Z and Y are jointly Gaussian. We have

$$\begin{aligned}\operatorname{Cov}(Z,Y) &= \mathbb{E}(BY - X - \mathbb{E}(BY - X))(Y - \mathbb{E}Y)^T \\ &= B\mathbb{E}(Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^T - \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)^T \\ &= B\operatorname{Cov}(Y) - \operatorname{Cov}(X,Y) = 0\end{aligned}$$

due to the choice of B. It follows from Lemma B.3 that Z and Y are independent. For any $g \in \mathcal{B}$, $g(Y)$ is $\sigma(Y)$-measurable, see e.g. Chung [18]. Therefore,

$$\mathbb{E}(Z - \mathbb{E}Z)^T g(Y) = \mathbb{E}(Z - \mathbb{E}Z)^T \mathbb{E}g(Y) = 0,$$

since $\mathbb{E}(Z - \mathbb{E}Z) = 0$. Lemma B.2 now yields the desired conclusion. □

Bibliography

[1] J. Albersmeyer. Effiziente Ableitungserzeugung in einem adaptiven BDF-Verfahren. Master's thesis, Universität Heidelberg, 2005.

[2] J. Albersmeyer, D. Beigel, C. Kirches, L. Wirsching, H. G. Bock, and J. P. Schlöder. Fast nonlinear model predictive control with an application in automotive engineering. In L. Magni, D.M. Raimondo, and F. Allgöwer, editors, *Lecture Notes in Control and Information Sciences*, volume 384, pages 471–480. Springer Verlag Berlin Heidelberg, 2009.

[3] J. Albersmeyer and H. G. Bock. Sensitivity generation in an adaptive BDF-method. In *Modeling, Simulation and Optimization of Complex Processes: Proceedings of the International Conference on High Performance Scientific Computing*, Hanoi, Vietnam, March 6-10, 2006.

[4] K. J. Åström. *Introduction to Stochastic Control Theory.* Academic Press, New York, London, 1970.

[5] M. Athans. The role and use of the stochastic Linear Quadratic Gaussian problem in control system design. *IEEE Transactions on Automatic Control*, VOL. AC-16, No. 6:529–552, 1971.

[6] R. R. Bahadur. Lectures on the theory of estimation. Institute of Mathematical Statistics, Beachwood, OH, 2002.

[7] R. J. Ballieu and K. Peiffer. Attractivity of the origin of the equation $\ddot{x} + f(t, x, \dot{x})|\dot{x}|^{\alpha}\dot{x} + g(x) = 0$. *Journal of Mathematical Analysis and Applications*, 65:321–332, 1978.

[8] Y. Bar-Shalom and E. Tse. Caution, probing, and the value of information in the control of uncertain systems. *Annals of Economic and Social Measurement*, 5/3:323–336, 1976.

[9] T. Basar. *Control Theory: Twenty-Five Seminal Papers (1932-1981).* Wiley-IEEE Press, 2001.

[10] D. P. Bertsekas. *Dynamic Programming and Optimal Control.* Athena Scientific, Belmont, Massachusetts, 1995.

[11] H. G. Bock. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen*, volume 183 of *Bonner Mathematische Schriften.* Universität Bonn, Bonn, 1987.

[12] H. G. Bock, M. Diehl, E. A. Kostina, and J. P. Schlöder. Constrained optimal feedback control of systems governed by large differential algebraic equations. In L. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders, editors, *Real-Time PDE-Constrained Optimization*, pages 3–24. SIAM, 2007.

[13] H. G. Bock and K. J. Plitt. A Multiple Shooting algorithm for direct solution of optimal control problems. In *Proceedings of the 9th IFAC World Congress*, volume IX, pages 242–247, Budapest, 1984. Pergamon Press.

[14] A. Bryson and Y. Ho. *Applied Optimal Control*. Hemisphere Publishing Corporation, 1975.

[15] W. K. Bühler. *Gauss: a Biographical Study*. Springer Verlag New York Inc., 1981.

[16] E. F. Camacho and C. Bordons. *Model Predictive Control*. Springer, 2nd edition, 2004.

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

[18] K. L. Chung. *A Course in Probability Theory*. Academic Press, 3rd edition, 2001.

[19] M. Diehl. *Real-Time Optimization for Large Scale Nonlinear Processes*. PhD thesis, Universität Heidelberg, 2001.

[20] M. Diehl, R. Amrit, and J. B. Rawlings. A Lyapunov function for economic optimizing Model Predictive Control. *IEEE Transactions on Automatic Control*, 56(3):703–707, 2011.

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

[22] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, 4th edition, 2010.

[23] S. Engell. Online optimizing control: The link between plant economics and process control. *Computer Aided Chemical Engineering*, 27:79–86, 2009.

[24] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 1996.

[25] L. C. Evans. An introduction to mathematical optimal control theory, version 0.2. `https://math.berkeley.edu/~evans/control.course.pdf`.

[26] A. A. Feldbaum. Dual control theory. I-IV. *Automation Remote Control*, Vol. 21,22, 1960-1961.

[27] N. Filatov and H. Unbehauen. *Adaptive Dual Control*. Lecture Notes in Control and Information Sciences. Springer Verlag, Berlin, 2004.

[28] R. Findeisen, L. Imsland, F. Allgöwer, and B. Foss. Output feedback stabilization of constrained systems with nonlinear predictive control. *Int. J. Robust Nonlinear Control*, 13:1211–1227, 2003.

[29] F. Fontes. A general framework to design stabilizing nonlinear model predictive controllers. *Systems & Control Letters*, 42:127–143, 2001.

[30] J. Frasch, L. Wirsching, S. Sager, and H. G. Bock. Mixed-level iteration schemes for nonlinear model predictive control. In *4th IFAC Nonlinear Model Predictive Control Conference*, 2012.

[31] J. V. Frasch, T. Kraus, W. Saeys, and M. Diehl. Moving horizon observation for autonomous operation of agricultural vehicles. *European Control Conference*, July 17-19, Zürich, Switzerland. 2013.

[32] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47:99–131, 2005.

[33] M. S. Grewal and A. P. Andrews. *Kalman Filtering: Theory and Practice Using MATLAB*. John Wiley & Sons, INC., 3rd edition, 2008.

[34] 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. *Automatic Control*, 50(5):546–558, 2005.

[35] L. Grüne. Economic receding horizon control without terminal constraints. *Automatica*, 49(3):725–734, 2013.

[36] L. Grüne and J. Pannek. *Nonlinear Model Predictive Control: Theory and Algorithms*. Springer-Verlag, London, 2011.

[37] T. A. N. Heirung, B. Foss, and B. E. Ydstie. MPC-based dual control with online experiment design. *Journal of Process Control*, 32:64–76, 2015.

[38] H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118(2):371–406, 2009.

[39] A. V. Ivanov. An asymptotic expansion for the distribution of the least squares estimator of the nonlinear regression parameter. *Theory of Probability and Its Applications*, XXI:557–570, 1976.

[40] A. V. Ivanov. *Asymptotic Theory of Nonlinear Regression*. Springer Science+Business Media Dordrecht, 1997.

[41] 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.

[42] A. Jadbabaie, J. Primbs, and J. Hauser. Unconstrained receding horizon control with no terminal cost. In *Proceedings of the American Control Conference*, volume 4, pages 3055–3060, Arlington, June 25–27, 2001.

[43] R. I. Jennrich. Asymptotic properties of nonlinear least squares estimators. *The Annals of Mathematical Statistics*, 40(3):633–643, 1969.

[44] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82:35–45, 1960.

[45] C. Kirches, L. Wirsching, S. Sager, and H. G. Bock. *Recent Advances in Optimization and its Applications in Engineering*, chapter Efficient Numerics for Nonlinear Model Predictive Control, pages 339–357. Springer Berlin Heidelberg, 2010.

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

[47] S. Körkel. Online experimental design for model validation. In *Rita Maria de Brito Alves, Claudio Augusto Oller do Nascimento, and Evaristo Chalbaud Biscaia Jr., editors, Proceedings of 10th International Symposium on Process Systems Engineering*, 2009.

[48] S. Körkel, E. A. 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(3-4):327–338, 2004.

[49] S. Körkel, A. Potschka, H.G. Bock, and S. Sager. A multiple shooting formulation for optimum experimental design. *Mathematical Programming (in revision)*.

[50] P. Kühl, M. Diehl, T. Kraus, J. P. Schlöder, and H. G. Bock. A real-time algorithm for moving horizon state and parameter estimation. *Computers and Chemical Engineering*, 35:71–83, 2011.

[51] D. Kundu. Asymptotic theory of least squares estimator of a particular nonlinear regression model. *Statistics & Probability Letters*, 18:13–17, 1993.

[52] H. C. La, A. Potschka, and H. Bock. Partial stability for nonlinear model predictive control. *(submitted)*.

[53] H. C. La, A. Potschka, J. P. Schlöder, and H. G. Bock. Dual control and information gain in controlling uncertain processes. In *Proceedings of the 11th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems. Trondheim, Norway, June* 6-8, 2016. *(accepted)*.

[54] H. C. La, A. Potschka, J. P. Schlöder, and H. G. Bock. Dual control and Online Optimal Experimental Design. *(submitted)*.

[55] H. C. La, J. P. Schlöder, and H. G. Bock. Structure of optimal samples in continuous nonlinear experimental design for parameter estimation. In *Proceedings of 6th International Conference on High Performance Scientific Computing. Hanoi, March* 16-20, 2015. *(accepted)*.

[56] J. P. LaSalle. Stability theory for ordinary differential equations. *Journal of Differential Equations*, 4:57–65, 1968.

[57] L. Ljung. *System Identification – Theory for the User*. Prentice Hall PRT, Upper Saddle River, N.J., 1999.

[58] S. Lucia and R. Paulen. Robust nonlinear model predictive control with reduction of uncertainty via robust optimal experiment design. In *Proc. of the 19th IFAC World Congress*, pages 1904–1909, Cape Town, 2014.

[59] E. Malinvaud. The consistency of nonlinear regressions. *The Annals of Mathematical Statistics*, 41:956–969, 1970.

[60] D. Q. Mayne and H. Michalska. Receding horizon control of nonlinear systems. *IEEE Transactions on Automatic Control*, 35:814–824, 1990.

[61] H. Michalska and D. Q. Mayne. Robust receding horizon control of constrained nonlinear systems. *IEEE Transactions on Automatic Control*, 38:1623–1633, 1993.

[62] P. Milgrom and I. Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, Vol. 70, No. 2:583–601, 2002.

[63] D. S. Moore and G. P. McCabe. *Introduction to the Practice of Statistics, 3rd ed.* W. H. Freeman and Company, New York, 2001.

[64] K. E. Morrison. Random walks with decreasing steps. Unpublished manuscript, California Polytechnic State University, 1998 (retrieved June 18, 2015 from `http://www.calpoly.edu/~kmorriso/Research/RandomWalks.pdf`).

[65] M. R. Osborne. On shooting method for boundary value problems. *Journal of Mathematical Analysis and Applications*, 27:417–433, 1969.

[66] A. Pázman. *Foundations of Optimum Experimental Design.* D. Reidel Publishing Company, 1986.

[67] A. Potschka. Handling path constraints in a direct multiple shooting method for optimal control problems. Master's thesis, Ruprecht-Karls-Universität Heidelberg, 2006.

[68] L. Pronzato and A. Pázman. *Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties*, volume 212 of *Lecture Notes in Statistics.* Springer, 2013.

[69] F. Pukelsheim. *Optimal Designs of Experiments.* John Wiley & Son, Inc., 1993.

[70] B. L. S. Prakasa Rao. On the exponential rate of convergence of the least squares estimator in the nonlinear regression model with Gaussian errors. *Statistics & Probability Letters*, 2:139–142, 1984.

[71] J. B. Rawlings and D. Q. Mayne. *Model Predictive Control: Theory and Design.* Nob Hill Pub, 2009.

[72] M. Reble and F. Allgöwer. Unconstrained model predictive control and suboptimality estimates for nonlinear continuous-time systems. *Automatica*, 48:1812–1817, 2012.

[73] J. A. Rice. *Mathematical Statistics and Data Analysis.* Thomson Brooks/Cole, 3rd edition, 2007.

[74] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis.* Springer, 1998.

[75] W. Rudin. *Principles of Mathematical Analysis, 3rd Edition.* McGraw-Hill, New York, 1976.

[76] W. Rudin. *Functional Analysis.* McGraw-Hill, 2nd edition, 1991.

[77] S. Sager. Sampling decisions in optimal experimental design in the light of Pontryagin's maximum principle. *SIAM J. Control Optim.*, 51(4):3181–3207, 2013.

[78] J. P. Schlöder. *Numerische Methoden zur Behandlung hochdimensionaler Aufgaben der Parameteridentifizierung.* PhD thesis, Hohe Mathematisch-Naturwissenschaftliche Fakultät der Rheinishen Friedrich-Wilhelms-Universität zu Bonn, 1987.

[79] G. A. F. Seber and C. J. Wild. *Nonlinear Regression.* Wiley, New York, 2003.

[80] H. W. Sorenson. *Parameter Estimation: Principles and Problems.* New York: M. Dekker, 1980.

[81] B. Srinivasan, S. Palanki, and D. Bonvin. Dynamic optimization of batch processes I. Characterization of the nominal solution. *Computer and Chemical Engineering*, 27:1–26, 2003.

[82] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis.* Springer Verlag, 2002.

[83] C. R. Vogel. *Computational Methods for Inverse Problems.* SIAM, 2002.

[84] V. I. Vorotnikov. *Partial Stability and Control.* Birkhäuser Boston, 1998.

[85] A. Walther and A. Griewank. Getting started with ADOL-C. In *U. Naumann und O. Schenk, Combinatorial Scientific Computing, Chapman-Hall CRC Computational Science*, pages 181–202, 2012.

[86] B. Wittenmark. Adaptive dual control methods: An overview. In *5th IFAC symposium on Adaptive Systems in Control and Signal Processing*, pages 67–72, 1995.

[87] V. M. Zavala and L. T. Biegler. The advanced-step NMPC controller: Optimality, stability and robustness. *Automatica*, 45:86–93, 2009.

[88] E. Zeidler. *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization.* Springer-Verlag New York Inc., 1985.