# Chapter 1
# Introduction to Nonlinear Model Predictive Control and Moving Horizon Estimation

Tor A. Johansen

**Abstract** Nonlinear model predictive control and moving horizon estimation are related methods since both are based on the concept of solving an optimization problem that involves a finite time horizon and a dynamic mathematical model. This chapter provides an introduction to these methods, with emphasis on how to formulate the optimization problem. Both theoretical and practical aspects are treated, ranging from theoretical concepts such as stability, existence and uniqueness of the solution, to practical challenges related to numerical optimization methods and computational complexity.

## 1.1 Introduction

The purpose of this chapter is to give an introduction to two of the most powerful tools that can be used to address nonlinear control and estimation problems - nonlinear model predictive control (NMPC) and nonlinear moving horizon estimation (NMHE). They are treated together since they are almost identical in approach and implementation - even though they solve two different and complementary problems.

The text is intended for advanced master and doctoral level students that have a solid background in linear and nonlinear control theory, and with a background in linear MPC, numerical methods for optimization and simulation, and state estimation using observers and the Extended Kalman Filter. Other excellent surveys to the topic and introductory texts can be found in Allgöwer et al (1999); Findeisen et al (2003b); Mayne et al (2000); Morari and Lee (1999); Rawlings (2000).

Tor A. Johansen
Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway, e-mail: tor.arne.johansen@itk.ntnu.no

### *1.1.1 Motivation and Main Ideas*

#### 1.1.1.1 Nonlinear Control

Consider the problem of controlling a multi-variable nonlinear system, subject to physical and operational constraints on the input and state. Well known systematic nonlinear control methods such as feedback linearization (Isidori (1989); Marino and Tomei (1995); Nijmeijer and van der Schaft (1990)) and constructive Lyapunov-based methods (Krstic et al (1995); Sepulchre et al (1997)) lead to very elegant solutions, but they depend on complicated design procedures that does not scale well to large systems and they are not developed in order to handle constraints in a systematic manner. The concept of optimal control, and in particular its practical implementation in terms of Nonlinear Model Predictive Control (NMPC) is an attractive alternative since the complexity of the control design and specification increases moderately with the size and complexity of the system. In particular for systems that can be adequately modeled with linear models, MPC has become the de-facto standard advanced control method in the process industries (Qin and Badgwell (1996)). This is due to its ability to handle large scale multivariable processes with tens or hundreds of inputs and states that must fulfill physical and operational constraints.

MPC involves the formulation and solution of a numerical optimization problem corresponding to a finite-horizon optimal control problem at each sampling instant. Since the state of the system is updated during each sampling period, a new optimization problem must be solved at each sampling interval. This is know as the receding horizon approach. With linear models the MPC problem is typically a quadratic or linear program, which is known to be convex and for which there exists a variety of numerical methods and software. While the numerical complexity of linear MPC may be a reasonable challenge with powerful computers being available, there is no doubt that NMPC is limited in its industrial impact due to the challenges of guaranteeing a global (or at least sufficiently good) solution to the resulting nonlinear optimization problem within the real-time requirements (Qin and Badgwell (2000)). Other limiting factors are the challenges of developing nonlinear dynamic models and state estimators. The nonlinear programming problem may have multiple local minima and will demand a much larger number of computations at each sample, even without providing any hard guarantees on the solution. Hence, NMPC is currently not a panacea that can be plugged in to solved any control problem. However, it is a powerful approach of great promise that has proven itself in several applications, Qin and Badgwell (2000); Foss and Schei (2007), and with further research in the direction of numerical implementation technology and modeling and state estimation methods, it may strengthen its position as the most powerful method available for certain classes of systems.

#### 1.1.1.2 Nonlinear Estimation

Consider the state estimation problem of nonlinear systems. A least-squares optimal state estimation problem can be formulated by minimizing a properly weighted least-squares criterion defined on the full data history horizon, subject to the nonlinear model equations, (Moraal and Grizzle (1995b); Rao et al (2003)). This is, however, impractical as infinite memory and processing will be needed as the amount of data grows unbounded with time. Alternatively, a well known sub-optimal estimator is given by an Extended Kalman Filter (EKF) which approximates this least-squares problem and defines a finite memory recursive algorithm suited for real-time implementation, where only the last measurement is used to update the state estimate, based on the past history being approximately summarized by estimates of the state and the error covariance matrix, Gelb (2002). Unfortunately, the EKF is based on various stochastic assumptions on noise and disturbances that are rarely met in practice, and in combination with nonlinearities and model uncertainty, this may lead to unacceptable performance of the EKF. A possible better use of the dynamic model and past history when updating the state estimate is made by a nonlinear Moving Horizon State Estimator (NMHE) that makes use of a finite memory moving window of both current and historical measurement data in the least-squares criterion, possibly in addition to known constraints on the state and uncertainty, and a state estimate and error covariance matrix estimate to estimate the arrival-cost at the beginning of the data window, see Rao et al (2003); Moraal and Grizzle (1995b); Alessandri et al (1999, 2008) for different formulation relying on somewhat different assumptions. Such an MHE can also be considered a sub-optimal approximation to an estimator that uses the full history of past data, and some empirical studies, Haseltine and Rawlings (2005) show that the NMHE can perform better than the EKF in terms of accuracy and robustness. It should also be mentioned that other variations of the Kalman filter, such as particle filters and the unscented Kalman filter, also show great promise for nonlinear state estimation (Rawlings and Bakshi (2006); Kandepu et al (2008); Bølviken et al (2001)) and are competitive alternatives to NMHE. Finally, we remark that nonlinear observers based on constructive Lyapunov design methods Krstic et al (1995); Sepulchre et al (1997) and nonlinear system theory (Marino and Tomei (1995); Isidori (1989)) are developed for certain classes of nonlinear systems and leads to very elegant and computationally efficient solutions, but are not easy to develop for large classes of high order multi-variable systems.

### *1.1.2 Historical Literature Review*

Originally, the MPC and MHE methods were developed fairly independently. More recently, with the development of algorithms for constrained NMPC and

NMHE their developments have converged and the methods are more often presented as duals of each other and with similar notation and terminology. One reason is the fundamental duality between estimation and control, Goodwin et al (2005), but equally important may be their dependence on nonlinear numerical optimization and similarities in the formulation of the optimization problems that leads to synergies when implementing practical solutions.

#### 1.1.2.1 Nonlinear Model Predictive Control

The nonlinear optimal control theory was developed in the 1950's and 1960's, resulting in powerful characterizations such as the maximum principle, Athans and Falb (1966) and dynamic programming, Bellman (1957). In the direct numerical optimal control literature, Hicks and Ray (1971); Deuflhard (1974); Biegler (1984); Bock and Plitt (1984); Betts (2001); Gill et al (1997); Bock et al (1999); von Stryk (1993), numerical methods to compute open loop control trajectories were central research topics. Problem formulations that included constraints on control and state variables were treated using numerical optimization.

NMPC involves the repetitive solution of an optimal control problem at each sampling instant in a receding horizon fashion. Unfortunately, there is no guarantee that the receding horizon implementation of a sequences of open loop optimal control solutions will perform well, or even be stable, when considering the closed loop system. This challenge, in combination with the tremendous success of *linear* MPC in the process industries, Qin and Badgwell (1996), lead to an increasing academic interest in NMPC research with focus on stability analysis and design modifications that guarantee stability and robustness. The early results Chen and Shaw (1982); Keerthi and Gilbert (1988); Mayne and Michalska (1990) boosted a large series of research, including Michalska and Mayne (1993); Alamir and Bornard (1995); Chen and Allgöwer (1998); Nicolao et al (2000); Scokaert et al (1999); Magni et al (2001a,b); Jadbabaie et al (2001); Mayne et al (2000). Industrial applications of NMPC have been reported, and are surveyed in Qin and Badgwell (2000); Foss and Schei (2007).

One of the early contributions of NMPC are given in Li and Biegler (1989), that uses linearization procedures and Gauss-Newton methods to provide a numerical procedure for NMPC based on SQP that makes only one Newton-iteration at each sampling instant. Theoretical results are also given in Li and Biegler (1990). The continuation/GMRES method of Ohtsuka (2004) is based on a similar philosophy of only one Newton-iteration per sample, while it is based on interior point methods. Recent NMPC research along similar ideas has benefited considerably from progress in numerical optimization, being able to take advantage of structural properties on the NMPC problem and general efficiency improvements, e.g. Biegler (2000); Diehl et al (2009); Tenny

et al (2004); Zavala and Biegler (2009), in addition to important issues such as robustness Magni et al (2003); Magni and Scattolini (2007); Limon et al (2006).

In parallel with the development of NMPC, researchers have developed so-called Real-Time Optimization (RTO) approaches, Sequeira et al (2002); Xiong and Jutan (2003). They are conceptually similar to NMPC, as they are generally based on nonlinear models (usually first principles models) and nonlinear programming. Their conceptual difference is that RTO uses static nonlinear models, while NMPC uses dynamic nonlinear models.

#### 1.1.2.2 Nonlinear Moving Horizon Estimation

Generalizing ideas from linear filtering, Jazwinski (1968), early formulations of NMHE were developed in Jang et al (1986); Ramamurthi et al (1993); Kim et al (1991); Tjoa and Biegler (1991); Glad (1983); Zimmer (1994); Michalska and Mayne (1995). A direct approach to the deterministic discrete-time nonlinear MHE problem is to view the problem as one of inverting a sequence of nonlinear algebraic equations defined from the state update and measurement equations, and some moving time horizon, Moraal and Grizzle (1995b).

Such discrete-time observers are formulated in the context of numerical nonlinear optimization and analyzed with respect to convergence in Rao et al (2003); Alessandri et al (1999, 2008); Raff et al (2005); Alamir (1999). In recent contributions, Biyik and Arcak (2006) provides results on how to use a continuous time model in the discrete time design, while issues related to parameterization are highlighted in Alamir (2007) computational efficiency are central targets of Zavala et al (2008); Alamir (2007); Alessandri et al (2008).

Uniform observability is a key assumption in most formulations and analysis of NMHE. For many practical problems, like combined state and parameter estimation problems, uniform observability is often not fulfilled and modifications are needed to achieve robustness, Moraal and Grizzle (1995a); Sui and Johansen (2010).

### *1.1.3 Notation*

Norms: For a vector $x \in \mathbb{R}^n$, let $||x|| = ||x||_2 = \sqrt{x^T x}$ denote the Euclidean norm, and $||x||_1 = |x_1| + ... + |x_N|$ and $||x||_\infty = \max_i |x_i|$. The weighted norms are for a given symmetric matrix $Q \succ 0$ given as $||x||_Q = \sqrt{x^T Q x}$ and we use the same notation also when $Q \succeq 0$. Vectors $x_1, x_2, ..., x_N$ are stacked into one large vector $x$ by the notation $x = \text{col}(x_1, x_2, ...., x_N)$.

For a continuous signal $x(t)$, where $t$ denotes continuous time, we let $x[t_0, t_1]$ denote the trajectory between $t_0 \leq t \leq t_1$.

### *1.1.4 Organization*

This chapter is organized in three main sections. In section 1.2 the formulation of NMPC optimization problems is described, focusing on the consequences of the various choices and challenges an engineer will face when designing and tuning an NMPC. Likewise, section 1.3 considers the formulation of NMHE optimization problems. The more detailed aspects of implementation in terms of numerical computations and solving the optimization problem, are treated on a general level common for both NMPC and NMHE, in section 1.4.

## 1.2 NMPC Optimization Problem Formulation

This section will focus on the *formulation* of the NMPC problem, while the detailed issues related to its *numerical solution* are postponed until section 1.4. It is, however, important to have in mind that these two issues are closely linked. While the NMPC problem formulation is driven by the specification of the control objective, constraints and dynamic model formulations, one should also consider potential numerical challenges at this point. In particular, important characteristics of the tradeoff between numerical accuracy and computational complexity are determined already at the point when the NMPC optimization problem is formulation through discretization, choice of parameterizations, and choice of decision variables and constraint formulations in the optimization problem. Some of these relationships are treated also in this section, together with fundamental properties of the optimization problem, including stability, convexity and the link between controllability and well-posedness of the optimization problem.

### *1.2.1 Continuous-time Model, Discretization and Finite Parameterization*

This section will introduce a basic nonlinear optimal control formulation starting from a continuous time model and a finite horizon where the objective is to minimize a cost function

$$J(u[0,T],x[0,T]) \triangleq \int_0^T \ell(x(t),u(t),t)dt + S(x(T),T) \tag{1.1}$$

$$\tag{1.2}$$

subject to the inequality constraints for all $t \in [0,T]$

$$u_{min} \ \leq \ u(t) \leq u_{max} \tag{1.3}$$

$$g(x(t),u(t),t) \leq 0 \tag{1.4}$$

and the evolution of the ordinary differential equation (ODE) given by

$$\frac{d}{dt}x(t) = f(x(t),u(t),t) \tag{1.5}$$

with given initial condition $x(0) \in \mathbb{R}^n$. The function $\ell$ is know as the stage cost, $S$ is the terminal cost, $T > 0$ is the horizon, and together these define the cost function $J$. The evolution of the state $x(t)$ is given by the function $f$ according to (1.5) and depends on the input signal $u(t) \in \mathbb{R}^m$ and time $t$, and forms an infinite-dimensional equality constraint to the optimal solution in the formulation above. In addition there is saturation on the input with minimum and maximum thresholds $u_{min}$ and $u_{max}$, respectively, and general inequality constraints jointly on states and inputs, point-wise in time $t \in [0,T]$, defined by the function $g$. These constraints may result from both physical and operational constraints of the control system and stability-preserving terminal sets that will be discussed later in section 1.2.3, see also Mayne et al (2000).The properties of $\ell$ and $S$ have consequences for the control performance, including stability, and must be carefully understood and tuned, Mayne et al (2000). We will return to this important issue in section 1.2.3. The explicit time-dependence in $f, g, \ell$ allows for time-varying reference trajectories, known disturbances and exogenous input signals to be accounted for in the optimal control problem formulation. Throughout this chapter we implicitly assume all the functions involved satisfy the necessary regularity assumptions, such as continuity and smoothness.

The above formulation basically defines an infinite-dimensional optimal control problem whose solution can be characterized using classical tools like calculus of variations, Pontryagin's maximum principle (Athans and Falb (1966)) and dynamic programming, Bellman (1957). In these *indirect methods* such characterizations of the solution can help us only in a very limited number of special cases to find an analytic exact representation of the solution. The most interesting and well known is the unconstrained linear quadratic regulator (LQR) where the feedback solution is a linear state feedback $u = Kx$ under additional assumptions on $T$ and $S$ that makes the cost function equivalent to an infinite horizon cost Athans and Falb (1966). More recently, explicit piecewise linear state feedback representation of the solution can be made for the linearly constrained LQR problem (Bemporad et al (2002)) and more generally for linearly constrained discrete-time piecewise

linear systems, Bemporad et al (2000), although the complexity of the exact representation may be prohibitive for anything but small scale systems.

Although numerical solutions can be found based on the characterizations of the indirect methods, In the context of NMPC we choose to restrict our attention to so-called *direct methods* that seems most promising and popular. They are characterized by discretization and finite parameterization being introduced in the optimal control problem formulation which is then directly solved with numerical methods. The principle of NMPC is to repeatedly solve finite-horizon optimal control problems of the above kind at each sampling instant. This means that the initial state $x(0)$ to (1.5) is viewed as the current state based on the most recent measurements, and the optimal control trajectory $u[0,T]$ solving the above problem is implemented for a short period of time (usually one sampling interval, typically much smaller than $T$) until the procedure is repeated and an updated optimal control trajectory is available. However, the solution of the above optimal control problem, requires reformulations for the following reasons

- The solution to the ordinary differential equation (1.5) with given initial conditions must generally be based on discretized to be handled by numerical integration since exact closed-form solutions of the ODE are usually not possible to formulate in the general nonlinear case. Viewed in a different way, the infinite number of equality constraints (1.5) must be represented by a finite approximation.
- The infinite-dimensional unknown solution $u[0,T]$ should be replaced by a finite number of decision variables to be able to define a finite-dimensional optimization problem that can be solved using numerical optimization.
- Measurements are typically sampled data available only at the sampling instants, such that an updated initial state $x(0)$ will normally be available only at defined sampling instants.
- Arbitrary control trajectories cannot be implemented since typically the control command can only be changed at defined sampling instants and is typically assumed to be constant (or some other simple sample-and-hold function such as linear) between the sampling instants.

In order to reformulate the problem into a finite-dimensional and practical setting, we will make the following assumptions that will allow the integral and differentiation operators to be approximated by numerical integration methods.

- The horizon $T$ is finite and given.
- The input signal $u[0,T]$ is assumed to be piecewise constant with a regular sampling interval $t_s$ such that $T$ is an integer multiple of $t_s$, and parameterized by a vector $U \in \mathbb{R}^p$ such that $u(t) = \mu(t,U) \in \mathbb{R}^r$ is piecewise continuous.
- An (approximate) solution to (1.5) is assumed to be defined in the form $x(t) = \phi(t,U,x(0))$ at $N$ discrete time instants $T_d = \{t_1, t_2, ..., t_N\} \subset [0,T]$ for some ODE solution function $\phi(\cdot)$. The discrete set of time instants $T_d$

results from discretization of the ODEs and its time instants may not be equidistant. A simulation of the ODEs embedded in the function $\phi(\cdot)$ may incorporate additional intermediate time-steps not included in $T_d$, since the purpose of $T_d$ is primarily to discretize the inequality constraints (1.3)-(1.4) at a finite number of representative points in time and to approximate the integral in (1.1) with a finite sum. In general, the time instants $T_d$ need not coincide with sampling instants.

The assumption of given horizon $T$ is typical for many NMPC problems, but there are important exceptions such as minimum-time formulations in e.g. robotics, Shin and McKay (1985), batch process control (Foss et al (1995); Nagy and Braatz (2003); Nagy et al (2007)), and other problems such as diving compression (Feng et al (2009)), where the horizon $T$ may be considered a free variable. The resulting modifications to the problem formulations may lead to additional challenges related to the time discretization and may make the optimization problem more challenging.

The basis for the NMPC is the nominal model (1.5), and we remark that model uncertainty, unknown disturbances and measurement errors are not accounted for in this formulation of the NMPC problem. Various extensions and variations that can relax many of the assumptions above can be made relatively easy as straightforward modifications to the basic problem formulation. Obviously, the ODEs (1.5) can result from the spatial discretization of a partial differential equation (PDE), and the problem formulation can be augmented with nonlinear algebraic constraints in a straightforward way to account for a differential-algebraic model (DAE) model formulation (Cervantes and Biegler (1998); Diehl et al (2002)). For simplicity of presentation, we stick to the formulation above and return to some alternatives and opportunities that will be discussed in later sections.

The parameterization of the input signal $\mu(t, U)$ on the horizon $t \in [0, T]$ is important and will influence both the control performance and computational performance. In general, it should satisfy the following objectives

- Be sufficiently flexible in order to allow for a solution of the reformulated optimal control problem close to the solution original problem (1.1)-(1.5).
- Be parsimonos in the sense that it does not contain unnecessary parameters that will lead to unnecessary computational complexity and numerical sensitivity.
- Be implementable within the capabilities of the control system hardware and software, meaning that particular consideration may be needed for any parameterization beyond a piecewise constant input trajectory that is restricted to change its value only at the sampling instants.

Based on the last very practical point, a general choice is the piecewise constant control input $\mu(t, U) = U_k$ for $t_k \leq t < t_{k+1}$ parameterized by the vector $U = col(U_0, ..., U_{N-1}) \in \mathbb{R}^{mN}$. Practical experience shows that the receding horizon implementation offers considerable flexibility for a NMPC to recover performance due to sub-optimality at each step. Consequently, it is

common practice to implement move-blocking strategies such that a smaller number of parameters is required by restricted the input from change at every sampling instant on the horizon, in particular towards the end of the horizon. For example, MPC has been successfully implemented for stable plants based on linear models by optimizing a constant input on the whole horizon, Qin and Badgwell (1996).

### *1.2.2 Numerical Optimal Control*

In this section the basic optimal control formulation in section 1.2.1 is reformulated into a form suitable for numeric solution by a nonlinear optimization solver.

As classified in Diehl et al (2009) there are two main avenues to direct numerical optimal control

- **The sequential approach.** The ODE constraint (1.5) is solved via numeric simulation when evaluating the cost and constraint functions. This means that the intermediate states $x(t_1), ...., x(t_N)$ disappear from the problem formulation by substitution into the cost and constraint functions, while the control trajectory parameters $U$ are treated as unknowns. This leads to a sequence of simulate-optimize iterations, often known as *Direct Single Shooting*, Hicks and Ray (1971); Sargent and Sullivan (1977); Kraft (1985).
- **The simultaneous approach.** The ODE constraints (1.5) are discretized in time and the resulting finite set of nonlinear algebraic equations are treated as nonlinear equality constraints. The intermediate states $x(t_1), ...., x(t_N)$ are treated as unknown variables together with the control trajectory parameters $U$, and the cost function is evaluated simply by replacing the integral (1.1) by a finite sum. This leads to simultaneous solution of the ODEs and the optimization problem with a larger number of constraints and variables. The most well known methods of this type are *Direct Multiple Shooting* (Deuflhard (1974); Bock and Plitt (1984); Bock et al (1999); Leineweber et al (2003)) and *Collocation methods*, (Tsang et al (1975); Biegler (1984); von Stryk (1993)).

It is fair to say that all the above mentioned approaches have advantages that could make them the method of choice when considering a specific problem. Already now we are in position to understand some of the differences

- The simultaneous approach involves a larger number of constraints and variables and therefore leads to "bigger problems". On the other hand, the cost and constraint function evaluation is much simpler and there are structural properties of the equations and numerical advantages that can be exploited in some cases. This will be discussed in section 1.4.

- Neglecting errors due to discretization and numerical approximations, all methods results in the same optimal control trajectory. Hence, one may expect the main difference between these alternatives to be related to numerical properties and computational complexity. Numerical accuracy of the solution is a consequence of discretization, round-off errors, sensitivity to initial conditions and input, differences in linear algebraic methods, etc. and must be balanced against computational cost. These aspects will be treated in more detail in section 1.4.
- Nonlinear optimization problems are generally non-convex, and the convergence and success of a given optimization algorithm depend largely on the initial guess provided for the solution. The sequential and simultaneous approach are in this sense fundamentally different, since the simultaneous approach not only requires an initial control trajectory guess, but also one for the state trajectory. The availability of a good initial guess for the state trajectory is an advantage that can be exploited by the simultaneous approach. On the other hand, the presence of nonlinear equality constraints (which by definition are non-convex) in the simultaneous approach, one cannot expect feasible initial guesses, which has consequences for the choice of numerical methods, and will be further discussed in section 1.4.
- The sequential approach may use more or less arbitrary and separate ODE and optimization solvers, which may in some cases be simple and convenient when compared to the simultaneous approach that tend to require more specialized and integrated numeric software combining these tasks. This may be a particularly important issue for industrial users that must use software tools based on an extensive set of requirements and constraints.

#### 1.2.2.1 Direct Single Shooting

In direct single shooting (Hicks and Ray (1971); Sargent and Sullivan (1977); Kraft (1985)), the ODE constraint (1.5) is eliminated by substituting its discretized numerical solution $x(t_k) = \phi(t_k, U, x(0))$ into the cost and constraints; minimize with respect to $U$ the cost

$$V^*(x(0)) = \min_{U \in \mathbb{R}^p} V(U; x(0)) \triangleq \sum_{k=1}^{N} \ell(\phi(t_k, U, x(0)), \mu(t_k, U), t_k)(t_k - t_{k-1}) + S(\phi(T, U, x(0)), T) \tag{1.6}$$

subject to

$$u_{min} \leq \mu(t_k, U) \leq u_{max}, \quad t_k \in T_d \tag{1.7}$$

$$g(\phi(t_k, U, x(0)), \mu(t_k, U), t_k) \leq 0, \quad t_k \in T_d \tag{1.8}$$

and the ODE solution function $\phi(\cdot)$ is the result of a numerical integration scheme. In its simplest form, an explicit integration scheme may be used

$$x(t_{k+1}) = F(x(t_k), \mu(t_k, U), t_k), \;\; x(t_0) = x(0) \text{ given}, \tag{1.9}$$

for $k = 0, ..., N-1$, leading to

$$\phi(t_k, U, x(0)) = F(\ldots F(F(x(0), \mu(t_0, U), t_0), \mu(t_1, U), t_1), ..., \mu(t_{k-1}, U), t_{k-1}) \tag{1.10}$$

However, $\phi(t_k, U, x(0))$ may also be computed using any other (implicit) discretization scheme in the simulation.

The problem (1.6) - (1.8) is a nonlinear program in $U$ parameterized by the initial state vector $x(0)$ and time. Dependence on time-varying external signals such as references and known disturbances are left implicit in order to keep the notation simple. The receding horizon MPC strategy will therefore re-optimize $U$ when new state or external input information appears, typically periodically at each sample. We assume the solution exists, and let it be denoted $U^*$.

We note that the introduction of common modifications such as terminal constraints and infeasibility relaxations still gives a nonlinear program, but with additional decision variables and constraints.

#### 1.2.2.2 Direct Collocation

In direct collocation (Tsang et al (1975); Biegler (1984); von Stryk (1993)) the numerical solution for $x(t_k)$ is not substituted into the cost and constraint functions, but the associated nonlinear algebraic equations resulting of an ODE discretization scheme are kept. Hence, the variables $x(t_k), \;\; k = 1, ...N$ are treated as unknown decision variables:

$$\begin{aligned} V^*(x(0)) &= \min_{U \in \mathbb{R}^p, x(t_1) \in \mathbb{R}^n, \ldots, x(t_N) \in \mathbb{R}^n} V(U, x(t_1), ..., x(t_N); x(0)) \\ &\triangleq \sum_{k=1}^{N} \ell(x(t_k), \mu(t_k, U), t_k)(t_k - t_{k-1}) + S(x(t_N), T) \end{aligned} \tag{1.11}$$

subject to

$$u_{min} \;\; \leq \mu(t_k, U) \leq u_{max}, \quad t_k \in T_d \tag{1.12}$$

$$g(x(t_k), \mu(t_k, U), t_k) \leq 0, \quad t_k \in T_d \tag{1.13}$$

$$F(x(t_{k+1}), x(t_k), \mu(t_k, U), t_k) = 0, \quad k = 0, ..., N-1 \tag{1.14}$$

$$x(t_0) = x(0) \text{ given} \tag{1.15}$$

where $F$ is a function defined by the discretization scheme of the ODE (1.5). We observe from (1.14) that it directly allows for implicit numerical integration methods to be used, and that the algebraic equations resulting from the implicit integration scheme will be solved simultaneously with the optimization.

The problem (1.11) - (1.13) is a nonlinear program in the variables $U, x(t_1), ..., x(t_N)$ parameterized by the initial state vector $x(0)$. In addition, dependence on time-varying external signals such as references and known disturbances are left implicit in order to keep the notation simple. The receding horizon MPC strategy will therefore re-optimize $U$ when new state or external input information appears, typically periodically at each sample. We assume the solution exists, and let it be denoted $U^*, x^*(t_1), ..., x^*(t_N)$.

#### 1.2.2.3 Direct Multiple Shooting

Direct multiple shooting (Deuflhard (1974); Bock and Plitt (1984); Bock et al (1999); Leineweber et al (2003)) combines elements of both direct single shooting and direct collocation. It is a simultaneous approach in the sense it reformulates the ODE (1.5) to a set of nonlinear algebraic equality constraints that are solved simultaneously with the optimization. It differs from the direct collocation method since an ODE solver is used to simulate the ODE (1.5) in each time interval $t_k \leq t \leq t_{k+1}$ for $k = 0, ..., N-1$:

$$\begin{aligned} V^*(x(0)) &= \min_{U \in \mathbb{R}^p, (x(t_1),...,x(t_N))^T \in \mathbb{R}^{nN}} V(U, x(t_1), ..., x(t_N); x(0)) \\ &\triangleq \sum_{k=1}^{N} \ell(x(t_k), \mu(t_k, U), t_k)(t_k - t_{k-1}) + S(x(t_N), T) \end{aligned} \quad (1.16)$$

subject to

$$u_{min} \leq \mu(t_k, U) \leq u_{max}, \quad t_k \in T_d \quad (1.17)$$

$$g(x(t_k), \mu(t_k, U), t_k) \leq 0, \quad t_k \in T_d \quad (1.18)$$

$$x(t_{k+1}) = \phi(x(t_k), \mu(t_k, U), t_k), \quad k = 0, ..., N-1 \quad (1.19)$$

$$x(t_0) = x(0) \text{ given}, \quad (1.20)$$

where $\phi$ is a function defined by the simulation of the ODE (1.5). The main difference between direct multiple shooting and direct collocation is due to the use of an arbitrary ODE solver between the time-instants in $T_d$. Direct multiple shooting may have advantages when adaptive discretization schemes are needed (due to stiff dynamics, for example) since they might require a varying number of grid points for each iteration of the solver. With multiple shooting this can in principle be "hidden" within the direct single shooting solver used between each time-instant in $T_d$, while it directly leads to a change

in the dimensions of the optimization problem at each iteration with a direct collocation method. Direct multiple shooting decouples the grids required for the point-wise discretization of the constraints (1.18) and the discretization grid required to integrated the ODE. In a sense, direct multiple shooting provides additional flexibility compared to both direct single shooting and direct collocation. On the other hand, direct collocation leads to a more sparse structure that can be exploited by the numerical optimization solver.

#### 1.2.2.4 The Nonlinear Program – Feasibility and Continuity

This section summarizes some features of the numeric optimization problem resulting from the direct approach to numerical optimal control in NMPC. Important issues related to the well-posedness of the problem are reviewed. They are related to existence and uniqueness of the solution and continuous dependence of the solution on data such as the initial state $x(0)$. These are again related to regularity properties and fundamental properties such as controllability.

In summary, all formulations reviewed in this section lead to a nonlinear optimization problem of the form

$$V^*(\theta) = \min_z V(z, \theta) \tag{1.21}$$

subject to

$$G(z, \theta) \leq 0 \tag{1.22}$$

$$H(z, \theta) = 0 \tag{1.23}$$

where $z$ is a vector of decision variables (control trajectory parameters, intermediate states, slack variables, etc.) while $\theta$ is a vector of parameters to the problem (initial states, parameters of reference trajectories, exogenous inputs, etc.).

Existence of a solution corresponds to feasibility of the optimization problem. We define the feasible set of parameters $\Theta_F$ as the set that contains all $\theta$ for which the optimization problem (1.21)-(1.23) has a solution $z^*(\theta)$

$$\Theta_F = \{z \mid \text{there exists a } z \text{ such that } G(z, \theta) \leq 0,\ H(z, \theta) = 0\} \tag{1.24}$$

The feasible set is a result of the dynamics of the systems and basically all design parameters of the NMPC problem. Generally speaking, it is desired to make this set as large as possible while fulfilling the physical and operational constraints of the control system. We will return to this design issue in section 1.2.3.

For simplicity, let us for the time being neglect the equality constraints (1.23). Using direct single shooting they can be eliminated and are thus not

important for the understanding of the fundamental issues in this section. For a given parameter $\theta_0 \in \Theta_F$, consider the Karush-Kuhn-Tucker (KKT) first-order necessary conditions for local optimality of (1.21)-(1.22); Nocedal and Wright (1999)

$$\nabla_z L(z_0;\theta_0) = 0 \tag{1.25}$$
$$G(z_0;\theta_0) \leq 0 \tag{1.26}$$
$$\mu_0 \geq 0 \tag{1.27}$$
$$\text{diag}(\mu_0)G(z_0;\theta_0) = 0 \tag{1.28}$$

are necessary for a local minimum $z_0$, with associated Lagrange multiplier $\mu_0$ and the Lagrangian defined as

$$L(z,\mu;\theta) \triangleq V(z;\theta) + \mu^T G(z;\theta) \tag{1.29}$$

Consider the optimal active set $\mathcal{A}_0$ at $\theta_0$, i.e. a set of indices to active constraints in (1.26). The above conditions are sufficient for local optimality of $z_0$ provided the following second order condition holds:

$$y^T \nabla_z^2 L(z_0,\mu_0;\theta_0) y > 0, \quad \text{for all } y \in \mathcal{F} - \{0\} \tag{1.30}$$

with $\mathcal{F}$ being the set of all directions where it is not clear from first order conditions if the cost will increase or decrease:

$$\mathcal{F} = \{y \mid \nabla_z G_{\mathcal{A}_0}(z_0;\theta_0)y \geq 0, \nabla_z G_i(z_0;\theta_0)y = 0, \text{ for all } i \text{ with } (\mu_0)_i > 0\} \tag{1.31}$$

The notation $G_{\mathcal{A}_0}$ means the rows of $G$ with indices in $\mathcal{A}_0$. The following result gives local regularity conditions for the optimal solution, Lagrange multipliers and optimal cost as functions of $\theta$.

**Assumption A1.** $V$ and $G$ are twice continuously differentiable in a neighborhood of $(z_0,\theta_0)$.

**Assumption A2.** The sufficient conditions (1.25)-(1.28) and (1.30) for a local minimum at $z_0$ hold.

**Assumption A3.** Linear independence constraint qualification (LICQ) holds, i.e. the active constraint gradients $\nabla_U G_{\mathcal{A}_0}(z_0;\theta_0)$ are linearly independent.

**Assumption A4.** Strict complementary slackness holds, i.e. $(\mu_0)_{\mathcal{A}_0} > 0$.

**Theorem 1.** *For a given $z_0$ and $\theta_0$ then under assumptions A1-A3, $z_0$ is a local isolated minimum, and for $\theta$ in a neighborhood of $\theta_0$, there exists a unique continuous function $z^*(\theta)$ satisfying $z^*(\theta_0) = z_0$ and the sufficient conditions for a local minimum.*

*If in addition A4 holds, then for $\theta$ in a neighborhood of $\theta_0$ the function $z^*(\theta)$ is differentiable and the associated Lagrange multipliers $\mu^*(\theta)$ exists, and are unique and continuously differentiable. Finally, the set of active con-*

*straints is unchanged, and the active constraint gradients are linearly independent at $z^*(\theta)$.*

The first part is proven in Kojima (1980), and the 2nd part follows from Theorem 3.2.2 in Fiacco (1983).

The concept of controllability of nonlinear systems can be defined in several ways. Here we have taken a pragmatic point of view, and focus on conditions that leads to feasibility of the solution, and continuity of the value function or solution as a function of the time-varying data $\theta$ that includes the initial conditions. In the context of numerical optimal control, issues related to lack of controllability or inappropriate design choices will typically manifest themselves in terms of infeasibility (no solution exists), indefiniteness of the Hessian (a global solution is not found), or singularity or poor conditioning of the Hessian (the solution is not unique and continuously dependent on the input data, or is highly sensitive to changes in decision variables). The latter case means that small changes in the state may require very large control actions to compensate. Since the above sufficient optimality conditions are practically impossible to verify a priori, these are important issues to be monitored by the practical NMPC algorithm based on output from the numerical solver in order to asses the quality of the NMPC design and identify problems related to lack of controllability or inappropriate design or tuning of the NMPC criterion and constraints.

The simplest special case for which strong properties can be guarantees a priori is the case of joint convexity:

**A5.** $V$ and $G$ are jointly convex for all $(z, \theta)$.

The optimal cost function can now be shown to have some regularity properties, Mangasarian and Rosen (1964):

**Theorem 2.** *Suppose A1-A5 holds. Then $X_F$ is a closed convex set, and $V^* : \Theta_F \to \mathbb{R}$ is convex and continuous.*

Convexity of $\Theta_F$ and $V^*$ is a direct consequence of A5, while continuity of $V^*$ can be established under weaker conditions; Fiacco (1983). We remark that $V^*$ is in general not differentiable, but properties such as local differentiability and directional differentiability can be investigated as shown in e.g. Fiacco (1983). Regularity properties of the solution function $z^*$ is a slightly more delicate issue, and essentially relies on stronger assumptions such as strict joint convexity that ensure uniqueness of the solution.

### *1.2.3 Tuning and Stability*

The specification of the NMPC control functionality and dynamic performance is essentially provided through the cost function and the constraints. We will not go into details on the practical tuning tradeoffs and the types of

physical and operational constraints, but note that one may typically choose $l_2$ or $l_1$ type cost function

$$\ell(x,u,t) = ||x - r_x(t))||_Q^2 + ||u - r_u(t)||_R^2 \tag{1.32}$$
$$\ell(x,u,t) = ||Q(x - r_x(t))||_1 + ||R(u - r_u(t)||_1 \tag{1.33}$$

where the properties of the weight matrices $Q \succeq 0$ and $R \succeq 0$ are essential for performance, and in some cases also stability. In the simplest case when there exists an $\varepsilon > 0$ such that

$$\ell(x,u,t) \geq \varepsilon(||x||^2 + ||u||^2) \tag{1.34}$$

it is clear that all states and control actions are directly observable through the cost function, an it follows intuitively that minimization of the cost function will influence all states that are controllable. Based on the similar arguments, it is in fact sufficient for stabilization that only the unstable modes of the system are observable through the cost function, such that $Q \succeq 0$ may be sufficient if weights are given on the unstable modes, Mayne et al (2000). In order to ensure uniqueness of the control trajectory it is generally recommended that $R \succ 0$. In summary, conventional LQR tuning guidelines (e.g. Athans and Falb (1966)) are very helpful as a starting point also for NMPC.

Although the effect of modeling errors and disturbances will be discussed in section 1.2.4.2, we remark that incorrect choice of the reference $r_u(t)$ for the control input may lead to a steady-state error that will be important to compensate for in many applications.

NMPC is based on the receding horizon control principle, where a finite horizon open loop optimal control problem solved at each sampling instant and the optimized control trajectory is implemented until a new optimized control trajectory is available at the next sampling instant. This leads to closed-loop control since each new optimized control trajectory is based on the most recent state information. However, the numerical optimal control problem solved at each sampling instant provides essentially an open-loop control trajectory. The finite-horizon cost function imposes in principle no stability requirement by itself, and with an unfortunate choice of design parameters (horizon $T$, weight matrices $Q$ and $R$, terminal cost $S$, and certain constraints) the closed loop NMPC may be unstable. In particular for open loop unstable systems, it is important to understand how these design parameters should be chosen to avoid an unstable NMPC.

#### 1.2.3.1 Stability Preserving Constraints And Cost-to-go

This section discusses stability of the NMPC in more depth, and how this property is related to design parameters in the cost function and constraints. The description will be fairly informal, and we avoid the technical details in

order to focus on the most important concepts. For simplicity we assumme that the objective is regulation to a constant set-point $r$. Further details and a more rigorous treatment of the topic are found in Chen and Allgöwer (1998); Mayne et al (2000); Michalska and Mayne (1993); Keerthi and Gilbert (1988); Mayne and Michalska (1990), and we remark that the concepts relevant for NMPC are essentially the same as for linear MPC.

The following principles are generally useful to ensure stability of an NMPC Mayne et al (2000):

- The control trajectory parameterization $\mu(t, U)$ must be "sufficiently rich" - most theoretical work assume piecewise constant control input trajectory that is allowed to move at each sampling instant.
- From the optimality principle of dynamic programming, Bellman (1957), an infinite horizon cost may be expected to have a stabilizing property. Theoretically, this leads to an infinite dimensional problem (except in simple special cases), so more practical approaches are
  - Sufficiently large horizon $T$. However, it is not obvious to know what is large enough, in particular for an open loop unstable system and when the constrained outputs are non-minimum phase (see Saberi et al (2002) for results on the importance of the zero-dynamics of the constrained outputs for the linear case).
  - A terminal cost chosen to approximate the cost-to-go, i.e. $S(x(T), T) \approx \int_{t=T}^{\infty} \ell(x(t), u(t), t)dt$ such that the total cost function approximates an infinite horizon cost. Unfortunately, the cost-to-go is generally hard to compute and simple approximations are usually chosen.
- Terminal set constraints of the type $x(t_N) \in \Omega$ that ensures that the state is regulated "close enough" to the set-point such that after $T$ it is a priori known that there exists a feasible and stabilizing controller that will ensure that $x(t), t \geq T$ never leaves $\Omega$ and eventually goes asymptotically to the set-point. There are many algorithms based on this philosophy, some of them are defined as dual mode NMPC (Michalska and Mayne (1993)) since they switch to a stabilizing simpler (non-NMPC) control law once $\Omega$ is reached, while others continue to use NMPC also in $\Omega$ with the confidence that there exist an (explicit or implicit) stabilizing control law that the NMPC may improve upon.
- Terminal equality constraints of the type $x(t_N) = r$, Keerthi and Gilbert (1988), that ensures convergence in finite time. This basically implies that the cost after time $T$ is zero, and is therefore related to both an infinite-cost strategy and a stability-preserving-constraint strategy.
- Finally, the idea of choosing the cost-to-go to approximate an infinite-horizon cost and the use of a terminal set may be combined. With the use of a terminal set it will be sufficient to approximate the cost-to-go for states that are within the terminal set, and simple tools like local linearization can be applied to make this a fairly practical approach; Chen and Allgöwer (1998).

A formal treatment of these issues are found in the references, see Mayne et al (2000) for additional references. The main tools are the use of either the value function $V^*(x)$ as a Lyapunov function, or investigating monotony of a sequences of value function values. Instead, we provide an example that is essentially similar to the method in Chen and Allgöwer (1998).

**Example.** Consider the discrete-time non-linear system

$$x(t_{k+1}) = F(x(t_k), u(t_k)) \tag{1.35}$$

where $x \in \mathbb{R}^n$ is the state, and $u \in \mathbb{R}^m$ is the input. We assume the control objective is regulation to the origin. For the current $x(t_k)$, we formulate the optimization problem

$$V^*(x(t_k)) = \min_U J(U, x(t_k)) \tag{1.36}$$

subject to $x_{k|k} = x(t_k)$ and

$$\begin{aligned} & y_{\min} \leq y_{k+i|k} \leq y_{\max}, \quad i = 1, ..., N \\ & u_{\min} \leq u_{k+i} \leq u_{\max}, \quad i = 0, 1, ..., N-1, \\ & x_{k+N|k} \in \Omega \\ & x_{k+i+1|k} = F(x_{k+i|k}, u_{k+i}), \quad i = 0, 1, ..., N-1 \\ & y_{k+i|k} = Cx_{k+i|k}, \quad i = 1, 2, ..., N \end{aligned} \tag{1.37}$$

with $U = \{u_k, u_{k+1}, ..., u_{k+N-1}\}$ and the cost function given by

$$J(U, x(t_k)) = \sum_{i=0}^{N-1} \left( ||x_{k+i|k}||_Q^2 + ||u_{k+i}||_R^2 \right) + ||x_{k+N|k}||_P^2 \tag{1.38}$$

The compact and convex terminal set $\Omega$ is defined by

$$\Omega = \{x \in \mathbb{R}^n \mid x^T P x \leq \alpha\} \tag{1.39}$$

where $P = P^T \succ 0$ and $\alpha > 0$ will be specified shortly. An optimal solution to the problem (1.36)-(1.37) is denoted $U^* = \{u_t^*, u_{t+1}^*, ..., u_{t+N-1}^*\}$, and the control input is chosen according to the receding horizon policy $u(t_k) = u_t^*$. This and similar optimization problems can be formulated in a concise form

$$V^*(x) = \min_U J(U, x) \quad \text{subject to} \quad G(U, x) \leq 0 \tag{1.40}$$

Define the set of $N$-step feasible initial states as follows

$$X_F = \{x \in \mathbb{R}^n \mid G(U, x) \leq 0 \text{ for some } U \in \mathbb{R}^{Nm}\} \tag{1.41}$$

Suppose $\Omega$ is a control invariant set, such that $X_F$ is a subset of the $N$-step stabilizable set, Kerrigan and Maciejowski (2000). Notice that the origin is an

equilibrium and interior point in $X_F$. It remains to specify $P \succ 0$ and $\alpha > 0$ such that $\Omega$ is a control invariant set. For this purpose, we use the ideas of Chen and Allgöwer (1998), where one simultaneously determine a linear feedback such that $\Omega$ is positively invariant under this feedback. Define the local linearization at the origin

$$A = \frac{\partial f}{\partial x}(0,0), \quad B = \frac{\partial F}{\partial u}(0,0) \tag{1.42}$$

Now, the following assumptions are made:

- $(A, B)$ is stabilizable.
- $P, Q, R \succ 0$.
- $y_{min} < 0 < y_{max}$ and $u_{min} < 0 < u_{max}$.
- The function $f$ is twice continuously differentiable, with $f(0,0) = 0$.

Since $(A, B)$ is stabilizable, let $K$ denote the associated LQ optimal gain matrix, such that $A_0 = A - BK$ is strictly Hurwitz. A discrete-time reformulation of Lemma 1 in Chen and Allgöwer (1998) can be made, Johansen (2004):

**Lemma 1.** *If $P \succ 0$ satisfies the Lyapunov-equation*

$$A_0^T P A_0 - P = -\kappa P - Q - K^T R K \tag{1.43}$$

*for some $\kappa > 0$, there exists a constant $\alpha > 0$ such that $\Omega$ defined in (1.39) satisfies*

1. $\Omega \subset \mathcal{C} = \{x \in \mathbb{R}^n \mid u_{min} \leq -Kx \leq u_{max}, y_{min} \leq Cx \leq y_{max}\}$.
2. *The autonomous nonlinear system*

$$x(t_{k+1}) = F(x(t_k), -Kx(t_k)) \tag{1.44}$$

   *is asymptotically stable for all $x(0) \in \Omega$, i.e. $\Omega$ is positively invariant.*
3. *The infinite-horizon cost for the system (1.44)*

$$J_\infty(x(t_k)) = \sum_{i=0}^{\infty} \left( ||x_{k+i|k}||_Q^2 + ||Kx_{k+i|k}||_R^2 \right) \tag{1.45}$$

   *satisfies $J_\infty(x) \leq x^T P x$ for all $x \in \Omega$.*

In order to prove this result we first remark that the Lyapunov-equation (1.43) is generally satisfied for sufficiently small $\kappa > 0$ because $A_0$ is strictly Hurwitz and the right-hand side is negative definite. One may define a set of the form

$$\Omega_{\alpha_1} = \{x \in \mathbb{R}^n \mid x^T P x \leq \alpha_1\} \tag{1.46}$$

with $\alpha_1 > 0$, such that $\Omega_{\alpha_1} \subseteq \mathcal{C}$, i.e. an ellipsoidal inner approximation $\Omega_{\alpha_1}$ to the polyhedron $\mathcal{C}$ where the input and state constraints are satisfied. Hence, the first claim holds for all $\alpha \in (0, \alpha_1]$.

Define the positive definite function $W(x) = x^T P x$. Along trajectories of the autonomous system (1.44) we have

$$\begin{aligned} W(x(t_{k+1})) - W(x(t_k)) &= \left(A_0 x(t_k) + \phi(x(t_k))\right)^T P \left(A_0 x(t_k) + \phi(x(t_k))\right) \\ &\quad - x^T(t_k) P x(t_k) \\ &= x^T(t_k) \left(A_0^T P A_0 - P\right) x(t_k) + \phi^T(x(t_k)) P \phi(x(t_k)) \\ &\quad + x^T(t_k)(A_0^T P + P A_0)\phi(x(t_k)) \end{aligned}$$

where $\phi(x) = F(x, -Kx) - A_0 x$ satisfies $\phi(0) = 0$. From (1.43)

$$\begin{aligned} W(x(t_{k+1})) - W(x(t_k)) &= -x^T(t_k) \left(Q + K^T R K + \kappa P\right) x(t_k) \\ &\quad + x^T(t_k)(A_0^T P + P A_0)\phi(x(t_k)) + \phi^T(x(t_k)) P \phi(x(t_k)) \end{aligned}$$

Let $L_\phi$ be a Lipschitz constant for $\phi$ in $\Omega_\alpha$ (which must exist because $f$ is differentiable). Since $\partial\phi/\partial x(0) = 0$ and $\phi$ is twice differentiable we can choose $L_\phi > 0$ as close to zero as desired by selecting $\alpha > 0$ sufficiently small. Hence, there exist $\alpha \in (0, \alpha_1]$ such that

$$W(x(t_{k+1})) - W(x(t_k)) \leq -x^T(t_k) \left(\frac{\kappa}{2} P + Q + K^T R K\right) x(t_k) \quad (1.47)$$

for all $x(t_k) \in \Omega$ and positive invariance of $\Omega$ follows since $\Omega$ is a level set of $W$.

Notice that from (1.47) we have

$$W(x(\infty)) - W(x(0)) \leq -J_\infty(x(0)) - \frac{\kappa}{2} \sum_{k=0}^{\infty} ||x(t_k)||_P^2 \quad (1.48)$$

and the third claim holds because $W(x(\infty)) = 0$ for all $x(0) \in \Omega$.

Hence, the result is proven, and it follows from Mayne et al (2000); Chen and Allgöwer (1998) that the RHC makes the origin asymptotically stable with region of attraction equal to the feasible set $X_F$. A procedure for selecting $P, \kappa$ and $\alpha$ can be adapted from Chen and Allgöwer (1998).

#### 1.2.3.2 Sub-optimal NMPC

It may be difficult to establish a non-conservative hard bound on the number of iterations required for convergence of the nonlinear programming problem that NMPC must solve numerically at each sampling instant. Furthermore, there may not be computational resources available to guarantee that a sufficient number of iterations can be computed, and only a local minimum may

be found. As an example, some NMPC methods will assume that only one iteration is performed per sample, Li and Biegler (1989, 1990). Hence, it is of interest to understand the consequences of not converting in terms of control performance loss. A fundamental result is given in Scokaert et al (1999), where it is shown that feasibility and descent (reduction in cost function compared to the control trajectory computed at the previous sample) is sufficient for asymptotic stability of NMPC provided that terminal constraints are included in the formulation. Hence, optimality is not required. In the same spirit, a computationally efficient and robust implementation of these ideas are pursued in Lazar et al (2008), and also exploited in the context of approximate NMPC Bemporad and Filippi (2003); Johansen (2004).

#### 1.2.3.3 Example: Compressor Surge Control

Consider the following 2nd-order compressor model Greitzer (1976); Gravdahl and Egeland (1997) with $x_1$ being normalized mass flow, $x_2$ normalized pressure and $u$ normalized mass flow through a close coupled valve in series with the compressor

$$\dot{x}_1 = B\left(\Psi_e(x_1) - x_2 - u\right) \tag{1.49}$$

$$\dot{x}_2 = \frac{1}{B}\left(x_1 - \Phi(x_2)\right) \tag{1.50}$$

The following compressor and valve characteristics are used

$$\Psi_e(x_1) = \psi_{c0} + H\left(1 + 1.5\left(\frac{x_1}{W} - 1\right) - 0.5\left(\frac{x_1}{W} - 1\right)^3\right)$$
$$\Phi(x_2) = \gamma \text{sign}(x_2)\sqrt{|x_2|}$$

with $\gamma = 0.5$, $B = 1$, $H = 0.18$, $\psi_{c0} = 0.3$ and $W = 0.25$. The control objective is to avoid surge, i.e. stabilize the system. This may be formulated as

$$\ell(x,u) = \alpha(x - x^*)^T(x - x^*) + \kappa u^2$$
$$S(x) = Rv^2 + \beta(x - x^*)^T(x - x^*)$$

with $\alpha, \beta, \kappa, \rho \geq 0$ and the set-point $x_1^* = 0.40$, $x_2^* = 0.60$ corresponds to an unstable equilibrium point. We have chosen $\alpha = 1$, $\beta = 0$, and $\kappa = 0.08$. The horizon is chosen as $T = 12$, which is split into $N = p = 15$ equal-sized intervals, using piecewise constant control input parameterization. Valve capacity requires the constraint

$$0 \leq u(t) \leq 0.3 \tag{1.51}$$

to hold, and the pressure constraint

$$x_2 \geq 0.4 - v \tag{1.52}$$

avoids operation too far left of the operating point. The variable $v \geq 0$ is a slack variable introduced in order to avoid infeasibility and $R = 8$ is its weight in the cost function.

A nonlinear optimization problem is formulated using direct single shooting where explicit Euler integration with step size 0.02 is applied to solve the ODE. Due to the unstable dynamics, this may not be the best choice, but it is sufficient for this simple example.

The NLP solution is shown in Figure 1.1 as a function $u^*(x)$. The corresponding optimal cost $V^*(x)$ is shown in Figure 1.2, and simulation results are shown in Figure 1.3, where the controller is switched on after $t = 20$. We note that it quickly stabilizes the deep surge oscillations.

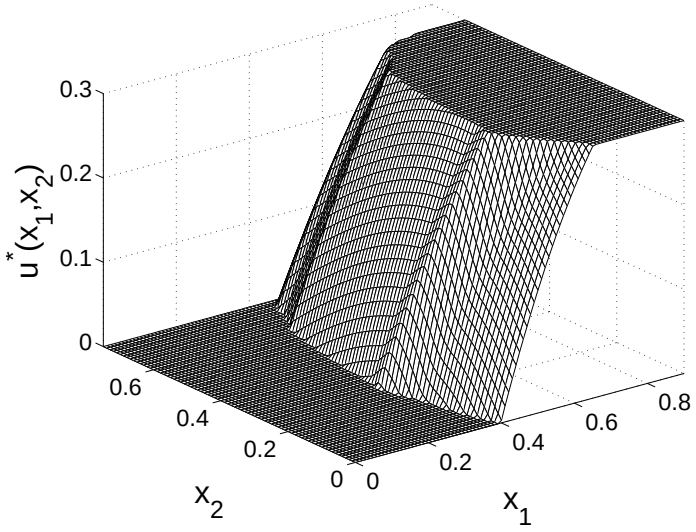


**Fig. 1.1** Feedback control law.

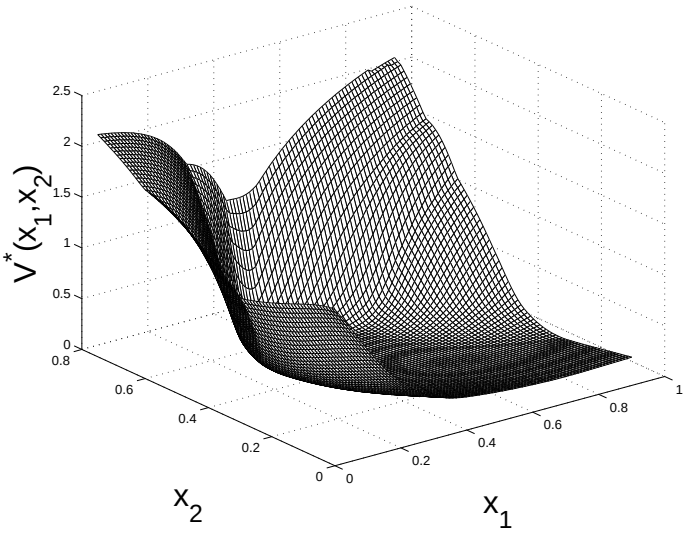


**Fig. 1.2** Optimal costs of the feedback control law.

### *1.2.4 Extensions and Variations of the Problem Formulation*

#### 1.2.4.1 Infeasibility Handling and Slack Variables

Feasibility of the NMPC optimization problem is an essential requirement for any meaningful state and reference command, and it is importance in practice that the NMPC optimization problem is formulated such that feasibility is ensured as far as possible by relaxing the constraints when needed and when possible. Obviously, physical constraints like input saturation can never be related, but operational constraints can generally be relaxed according to certain priorities under the additional requirement that safety constraints are fulfilled by a separate system (like an emergency shutdown system, pressure relief valves, or by functions in a decentralized control system). Stability-enforcing terminal constraints may also be relaxed in practice, or even skipped completely, since they tend to be conservative and often not needed when the NMPC is otherwise carefully designed, in particular for open loop stable systems.

A general way to reformulate an optimization problem to guarantee feasibility is to use slack variables (e.g. Vada et al (1999)). Taking the fairly general NLP formulation (1.21)-(1.23) as the starting point, we reformulate it in the following way

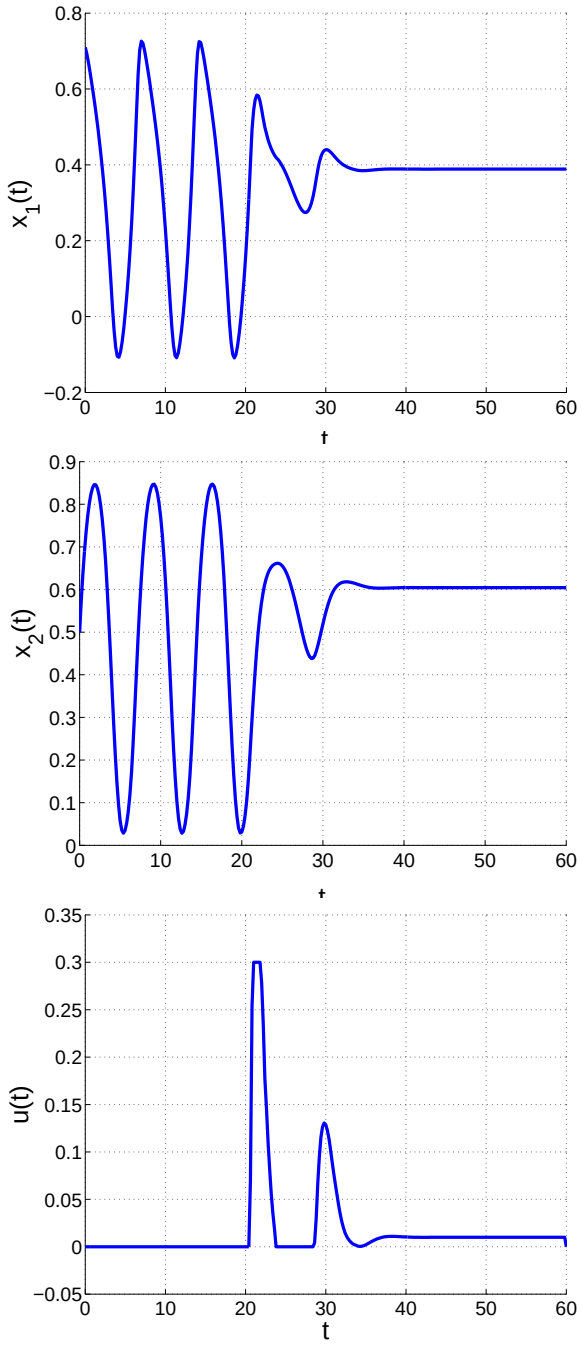


**Fig. 1.3** Simulation of compressor with nonlinear MPC.

$$V_s^*(\theta) = \min_{z,s,q} V(z,\theta) + ||W_s s||_1 + ||W_q q||_1 \tag{1.53}$$

subject to

$$G(z,\theta) \leq s \tag{1.54}$$
$$H(z,\theta) = q \tag{1.55}$$
$$s \geq 0 \tag{1.56}$$

where $W_s \succeq 0$ and $W_q \succeq 0$ are weight matrices of appropriate dimension. They are usually chosen such that the two latter penalty terms of (1.53) dominates the first term in order to ensure that the feasibility constraints are not relaxed when not needed.

#### 1.2.4.2 Robustness

Practical industrial experience shows that MPC tend to be inherently robust, Qin and Badgwell (1996, 2000), even without any particular consideration in the design phase beyond ensuring the accuracy of dynamic models and formulating realistic specifications in terms of operational constraints and cost function weights. In addition, mechanisms to handle steady state model errors (integral action like mechanisms) are usually implemented.

As a contrast to this practical experience, it is shown by examples, Grimm et al (2004), that when the NMPC problem involves state constraints, or terminal constraints in combination with short prediction horizons, the asymptotic stability of the closed-loop may have not be robust. A necessary condition for lack of robustness is that the value function and state feedback law are discontinuous, Grimm et al (2004), while at the same time lack of continuity does not necessarily lead to lack of robustness, Lazar et al (2007).

There exist a wide range of NMPC formulation that include robustness into the formulation of the optimization problem. One can mainly distinguish between three types of approaches; stochastic NMPC, min-max NMPC, and mechanisms to avoid steady-state errors.

There are two formulations of min-max NMPC: the open-loop and the closed-loop formulation (see Magni and Scattolini (2007) for review of the min-max NMPC approaches). The open-loop min-max NMPC (Michalska and Mayne (1993); Limon et al (2002); Magni and Scattolini (2007)) guarantees the robust stability and the robust feasibility of the system, but it may be very conservative since the control sequence has to ensure constraints fulfillment for all possible uncertainty scenarios without considering the fact that future measurements of the state contain information about past uncertainty values. As a result, the open-loop min-max NMPC controllers may have a small feasible set and a poor performance because they do not include the effect of feedback provided by the receding horizon strategy of MPC.

Most min-max MPC robustness approaches assume a fairly simple additive uncertainty model of the form

$$x_{k+1} = F(x_k, u_k) + w_k \tag{1.57}$$

where some bound on the unknown uncertainty $w_k$ is assumed. The conservativeness of the open-loop approaches is overcome by the closed-loop min-max NMPC (Magni et al (2003); Magni and Scattolini (2007); Limon et al (2006)), where the optimization is performed over a sequence of feedback control policies. With the closed-loop approach, the min-max NMPC problem represents a differential game where the controller is the minimizing player and the disturbance is the output of the maximizing player. The controller chooses the control input as a function of the current state so as to ensure that the effect of the disturbance on the system output is sufficiently small for any choice made by the maximizing player. In this way, the closed-loop min-max NMPC would guarantee a larger feasible set and a higher level of performance compared to the open-loop min-max NMPC (Magni et al (2003)).

Stochastic NMPC formulations are based on a probabilistic description of uncertainty, and can also be characterized as open-loop Cannon et al (2009); Kantas et al (2009) and closed-loop Goodwin et al (2009); Arellano-Garcia et al (2007) similarly to min-max robust NMPC as described above. They also share similar challenges due to significantly increased computational complexity when compared to nominal NMPC formulations.

The reformulation of nonlinear models as Linear Parameter Varying (LPV) models allows for the use of linear and bi-linear matrix inequality formulations of robust NMPC, Angeli et al (2000); Casavola et al (2003); Wan and Kothare (2004). The embedding of nonlinear systems into the class of LPV models

$$x_{k+1} = A(p_k)x_k + B(p_k)u_k + w(p_k) \tag{1.58}$$

leads to loss of information in the model that leads to more conservative robust control. However, using tools of semi-definite and convex programming, Boyd et al (1994), the LPV re-formulation allows for the computational complexity to be significantly reduced in many cases. In (1.58), $p_k$ is a parameter whose value is known to belong to some bounded set, and some approaches also assume that its time-derivative has a known bound, and the LPV re-formulation clearly allows a richer class of uncertainty to be modeled, compared to (1.57).

Steady-state control errors may result if there are steady-state model errors. While linear control design offers several tools to deal with this problem (including integral action, integrating models in linear MPC, and others), not all of them are directly transferable to nonlinear systems. The commonly used cure for steady-state errors in MPC, which can be directly transferred to NMPC, appears to be the use of a state estimator or observer that estimates an input or output disturbance for direct compensation in the NMPC

cost function, Muske and Badgwell (2002); Pannocchia and Rawlings (2003); Pannocchia and Bemporad (2007); Borrelli and Morari (2007).

#### 1.2.4.3 Observers and Output Feedback

Most formulations of nonlinear MPC assume state feedback. They are usually based on state space models, e.g. Balchen et al (1992); Foss and Schei (2007), although certain black-box using discrete-time nonlinear input/output models have also been proposed Nørgaard et al (2000); Åksesson and Toivonen (2006). Since all states are usually not measured, any implementation of NMPC based on a state space model will require a state estimator, which is often a critical component of an NMPC Kolås et al (2008). State space models have the advantage that they are most conveniently based on first principles.

Although practical rules of thumb for observer design such as separation of time-scales (typically one order of magnitude faster state estimator relative to the control loop response time) tend to be applicable in practical implementations also for NMPC, there also exist a number of rigorous theoretical results on the stability of the combination of observers with NMPC, see Findeisen et al (2003b) for an overview. Although a general separation principles does not exists for NMPC, there are some results in this direction, Findeisen et al (2003a); Adetola and Guay (2003); Messina et al (2005); Roset et al (2006).

#### 1.2.4.4 Mixed-integer MPC

General NMPC formulations based on nonlinear models suffer from the fact that it is hard to verify whether the underlying optimization problem is convex or not, such that in general it must be assumed to be non-convex. At the same time, all practical optimization solvers will assume some form of local convexity and guarantee convergence only to good initial guesses for the solution. This challenge will be further discussed in section 1.4. On the other hand, NMPC based on piecewise linear (PWL) models and cost functions will in general lead to mixed-integer linear programs (MI-LP) for which there exists solvers that guarantee global convergence, Tyler and Morari (1999); Bemporad and Morari (1999). The equivalence between a wide class of hybrid systems models, mixed logic models and PWL models, Heemels et al (2001), makes this approach attractive in many practical applications. Despite its applicability and importance, we only remark that the MI-LP theory and software are well developed, and refer to the references above and the large literature on MI-LP, Williams (1999).

#### 1.2.4.5 Decentralized and Distributed NMPC

Recently, several approaches for decentralized and distributed implementation of NMPC algorithms have been developed. A review of architectures for distributed and hierarchical MPC can be found in Scattolini (2009). The possibility to use MPC in a decentralized fashion has the advantage to reduce the original, large size, optimization problem into a number of smaller and more tractable ones.

In Magni and Scattolini (2006), a stabilizing decentralized MPC algorithm for nonlinear systems consisting of several interconnected local subsystems is developed. It is derived under the main assumptions that no information can be exchanged between local control laws, i.e. the coupling between the subsystems is ignored, and only input constraints are imposed on the system. In Dunbar and Murray (2006), it is supposed that the dynamics and constraints of the nonlinear subsystems are decoupled, but their state vectors are coupled in a single cost function of a finite horizon optimal control problem. In Keviczky et al (2006), an optimal control problem for a set of dynamically decoupled nonlinear systems, where the cost function and constraints couple the dynamical behavior of the systems, is solved.

## 1.3 NMHE Optimization Problem Formulation

In this section we consider the formulation of the NMHE optimization problem, and we follow as similar organization as section 1.2, with focus on the formulation of the optimization problem and the link between fundamental properties such as observability, detectability and existence and uniqueness of the solution.

### *1.3.1 Basic Problem Formulation*

The state estimation problem is to determine the current state based on a sequence of past and current measurements at discrete time instants, and the use of a dynamic model. For simplicity, we will assume data are available via synchronous sampling. Extension to be more general situation when data from the different sensors and data channels are asynchronous are conceptually straightforward and does not lead to any fundamental complications, but the mathematical notation requires many more indices and becomes unnecessarily tedious for an introduction to the topic. The problem can be treated by careful discretization of the continuous-time system to take asynchronous data into account, or a more pragmatic approach would be to rely on digital signal processing technique of interpolation and extrapolation for

pre-processing the data before used in the NMHE in order to artificially provide synchronized data as required at each sampling instant, Proakis and Manolakis (1996).

At the time $t_k$ corresponding to the discrete time index $k$ we consider a set of $N+1$ sampling instants $T_s = \{t_{k-N}, t_{k-N+1}, ...., t_k\}$, where the following synchronized window of output and input data are available

$$Y_k = \text{col}(y(t_{k-N}), y(t_{k-N+1}), ....., y(t_k))$$
$$U_k = \text{col}(u(t_{k-N}), u(t_{k-N+1}), ....., u(t_k))$$

where $y(t) \in \mathbb{R}^r$ and $u(t) \in \mathbb{R}^m$. We assume without loss of generality that sampling is periodic, i.e. the horizon $T = t_k - t_{k-N}$ and the sampling interval $t_s = t_i - t_{i-1}$ are constant. The inputs and outputs may be related by an ODE model

$$\frac{d}{dt}x(t) = f(x(t), u(t), w(t)) \tag{1.59a}$$
$$y(t) = h(x(t), u(t)) + v(t) \tag{1.59b}$$

with unknown initial condition $x(t_{k-N}) \subset \mathbb{R}^n$. The variable $w$ includes unknown model errors and disturbances, and $v$ includes unknown additive measurement errors. In addition, one may have available a priori information about $x(t)$ in the form of constraints on states and uncertainty

$$\text{col}(x(t), w(t), v(t)) \in X \times W \times V, \quad t \in [t_{k-N}, t_k] \tag{1.60}$$

for some compact sets $X, W$ and $V$. The constraints may result from operational knowledge of the system, or physical properties of the states (such as chemical composition never being negative at any point in time). More generally, such a priori knowledge may incorporate more complex statements that motivates a more general constraint formulation

$$C(x(t), w(t), v(t), t) \leq 0, \quad t \in [t_{k-N}, t_k] \tag{1.61}$$

The above constraint could incorporate time-varying information and statements that involves the interaction between two ore more variables - for example that a gas pressure is always below a certain threshold, expressed through the product of gas mass and temperature through the ideal gas law. One may also have a priori information that is not linked to a particular time instant, like that the average value of a certain variable is known to stay between certain upper and lower bounds or that the measurement noise has zero mean, which can be expressed as

$$\int_{t_{k-N}}^{t_k} c(x(t), w(t), v(t))dt \leq 0 \tag{1.62}$$

The state estimation problem is essentially to estimate $x(t_k)$ based on the $N+1$ data samples, the model, and the a priori information given in the form of constraints.

#### 1.3.1.1 Observability

The concept of observability is essential in order to formulate and understand the NMHE problem. In this section we will for convenience assume that the dynamic model system (1.59) is discretized in the form of a state space formulation

$$x_{k+1} = F(x_k, u_k, w_k) \tag{1.63a}$$

$$y_k = h(x_k, u_k) + v_k \tag{1.63b}$$

with the convenient notation $u_k = u(t_k), y_k = y(t_k), v_k = v(t_k), w_k = w(t_k)$. In this section, we will neglect the constraints (1.60)-(1.62) since they are not important for the observability concept. Furthermore, the process noise $v_k$ and measurement noise $w_k$ will also be set to zero and neglected in this section when defining the concept of observability. Note that by using the discrete-time equation (1.63a) recursively with initial condition $x(t_{k-N})$ and $v_k = 0$ and $w_k = 0$, one will uniquely determine $x(t), t \geq t_{k-N}$, including the current state $x(t_k)$ that we want to estimate.

To express $Y_k$ as a function of $x_{k-N}$ and $U_k$ under these conditions, denote $F_k(x_k) = F(x_k, u_k, 0)$ and $h_k(x_k) = h(x_k, u_k)$, and note from (1.63b) that the following algebraic map can be formulated, Moraal and Grizzle (1995b):

$$Y_k = H(x_{k-N}, U_k) = \begin{bmatrix} h^{u_{k-N}}(x_{k-N}) \\ h^{u_{k-N+1}} \circ F_{k-N}(x_{k-N}) \\ \vdots \\ h^{u_k} \circ F_{k-1} \circ \cdots \circ F_{k-N}(x_{k-N}) \end{bmatrix} \tag{1.64}$$

Hence, without the presence of any uncertainty and constraints, the state estimation problem is equivalent to the inversion of this set of nonlinear algebraic equations, like in the case of a linear system when full rank of the observability matrix is equivalent to observability. In order to better understand the similarities between the linear and nonlinear case, consider the linear system $x_{k+1} = Ax_k + Bu_k$ with output $y_k = Cx_k$. The (1.64) corresponds to

$$Y_k = \mathbb{C}_N x_{k-N} + \mathbb{B}_N U_k \tag{1.65}$$

where the matrix $\mathbb{C}_N$ is defined by

$$\mathbb{C}_N = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^N \end{pmatrix} \tag{1.66}$$

and $\mathbb{B}_N$ is a matrix that contains blocks of the form $CA^iB$. Clearly, the state can be uniquely determined from the window of past inputs and outputs by inverting the linear algebraic equations (1.65) if and only if $\mathbb{C}_N$ has full rank. It is well known from linear systems theory that $\text{rank}(\mathbb{C}_N) = \text{rank}(\mathbb{C}_n)$ for $N \geq n$, where $\mathbb{C}_n$ is known as the observability matrix. Similarly, in the nonlinear case, conditions that ensure that the inverse problem is well-posed (Tikhonov and Arsenin (1977)) in the sense that the inverse of (1.64) exists, is unique, and depends continuously on the data $U_k$ and $Y_k$ are of fundamental importance and essentially amounts to the concept of observability.

**Definition 1 (Moraal and Grizzle (1995b)).** The system (1.63) is *$N$-observable* if there exists a $K$-function $\varphi$ such that for all $x_1, x_2 \in \mathbb{X}$ there exists a feasible $U_k \in \mathbb{U}^{N+1}$ such that

$$\varphi(||x_1 - x_2||^2) \leq ||H(x_1, U_k) - H(x_2, U_k)||^2.$$

**Definition 2 (Sui and Johansen (2010)).** The input $U_k \in \mathbb{U}^{N+1}$ is said to be *$N$-exciting* for the $N$-observable system (1.63) at time index $k$ if there exists a $K$-function $\varphi_k$ that for all $x_1, x_2 \in \mathbb{X}$ satisfies

$$\varphi_t(||x_1 - x_2||^2) \leq ||H(x_1, U_k) - H(x_2, U_k)||^2.$$

From Proposition 2.4.7 in Abraham et al (1983), we have

$$H(x_1, U_k) - H(x_2, U_k) = \Phi_k(x_1, x_2)(x_1 - x_2), \tag{1.67}$$

where

$$\Phi_k(x_1, x_2) = \int_0^1 \frac{\partial}{\partial x} H((1-s)x_2 + sx_1, U_k) ds. \tag{1.68}$$

An observability rank condition can be formulated similar to the linear case outlined above (see also Moraal and Grizzle (1995b); Alessandri et al (2008); Fiacco (1983) and others for similar results):

**Lemma 2.** *If $\mathbb{X}$ and $\mathbb{U}$ are compact and convex sets, the functions $F$ and $h$ are twice differentiable on $\mathbb{X} \times \mathbb{U}$ and the Jacobian matrix $\frac{\partial H}{\partial x}(x, U_k)$ has full rank (equal to $n$) for all $x \in \mathbb{X}$ and some $U_k \in \mathbb{U}^{N+1}$, then the system is $N$-observable and the input $U_k$ is $N$-exciting for the system (1.63) at time index $k$.*

*Proof (Sui and Johansen (2010)).* Due to the observability rank condition being satisfied, $\Phi_k^T(\cdot)\Phi_k(\cdot) > 0$ and the system of nonlinear algebraic equations (1.67) can be inverted as follows:

$$x_1 - x_2 = \Phi_k^+(x_1, x_2)\big(H(x_1, U_k) - H(x_2, U_k)\big),$$
$$\Rightarrow \frac{1}{\pi_k^2(x_1, x_2)}\|x_1 - x_2\|^2 \leq \big\|H(x_1, U_k) - H(x_2, U_k)\big\|^2,$$

where $\pi_k(x_1, x_2) = \|\Phi_k^+(x_1, x_2)\|$. This proves that the conditions in Definitions 1 and 2 hold with $\varphi(s) = s/\overline{p}^2$ where

$$\overline{p} = \sup_{x_1, x_2 \in \mathbb{X}, U_k \in \mathbb{U}^{N+1}} ||\Phi_k^+(x_1, x_2)|| \tag{1.69}$$

is bounded due to $F$ and $h$ are twice differentiable on the compact set $\mathbb{X} \times \mathbb{U}$. □

This condition is a natural generalization of the linear observability matrix rank condition since

$$\frac{\partial H}{\partial x}(x, U_k) = \mathbb{C}_N \tag{1.70}$$

for a linear system, and the full rank condition of $\mathbb{C}_n$ is completely equivalent to observability for $N \geq n$. A fundamental difference is that in the nonlinear case the rank of the matrix $\frac{\partial H}{\partial x}(x, U_k)$ depends on both the current state $x$ and the current and past inputs $U_k$. This means that in the nonlinear case, successful asymptotic state estimation may depend on state and input trajectories, in strong contrast to the linear case where only the initial state influences the transient behavior of the observer (neglecting the influence of noise and disturbances in this discussion).

The role of the horizon parameter $N$ can also be understood from the above discussion. While $N = n$ is generally sufficient for an estimate to be computable for observable linear systems, the benefits of choosing $N$ larger is two-fold: The input data $U_k$ may be $N$-exciting for a nonlinear system for sufficiently large $N$, but not for $N = n$, and second, a larger $N$ will improve robustness to noise and uncertainty via a filtering effect. The possible disadvantages of choosing $N$ very large are increased computational complexity and too much filtering leading to slow convergence of the estimates.

Define the $N$-information vector at time index $k$ as

$$I_k = \text{col}(y_{k-N}, \ldots, y_k, u_{k-N}, \ldots, u_k).$$

When a system is not $N$-observable, it is not possible to reconstruct exactly all the state components from the $N$-information vector. However, in some cases one may be able to reconstruct exactly at least some components, based on the $N$-information vector, and the remaining components can be reconstructed asymptotically. This corresponds to the notion of detectability, where we suppose there exists a coordinate transform $\mathbb{T} : \mathbb{X} \to \mathbb{D} \subseteq R^n$, where $\mathbb{D}$ is the convex hull of $\mathbb{T}(\mathbb{X})$:

$$d = \text{col}(\xi, z) = \mathbb{T}(x) \tag{1.71}$$

such that the following dynamics are equivalent to (1.63) for any initial condition in $\mathbb{X}$ and inputs in $\mathbb{U}$,

$$\xi_{k+1} = F_1(\xi_k, z_k, u_k) \tag{1.72a}$$
$$z_{k+1} = F_2(z_k, u_k) \tag{1.72b}$$
$$y_k = g(z_k, u_k). \tag{1.72c}$$

This transform effectively partitions the state $x$ into an observable state $z$ and an unobservable state $\xi$. The following strong detectability definition is taken from Moraal and Grizzle (1995a):

**Definition 3.** The system (1.63) is *strongly $N$-detectable* if
(1) there exists a coordinate transform $\mathbb{T} : \mathbb{X} \to \mathbb{D}$ that brings the system in the form (1.72);
(2) the sub-system (1.72b)-(1.72c) is $N$-observable;
(3) the sub-system (1.72a) has uniformly contractive dynamics, i.e. there exists a constant $L_1 < 1$ such that for all $\mathrm{col}(\xi_1, z) \in \mathbb{D}, \mathrm{col}(\xi_2, z) \in \mathbb{D}$ and $u \in \mathbb{U}$, the function $F_1$ satisfies

$$||F_1(\xi_1, z, u) - F_1(\xi_2, z, u)||' \leq L_1 ||\xi_1 - \xi_2||'. \tag{1.73}$$

with a suitable norm $||\cdot||'$.

It is remarked that since there is considerable freedom in the choice of transform $\mathbb{T}$ and the norm $||\cdot||'$, the contraction assumption in part 3 of the definition is not very restrictive. For linear systems, it is equivalent to the conventional detectability definition.

**Definition 4.** The input $U_k$ is said to be *$N$-exciting* for a strongly $N$-detectable system (1.63) at time index $k$ if it is $N$-exciting for the sub-system (1.72b)-(1.72c) at time index $k$.

If the input $U_t$ is not $N$-exciting at certain points in time, the state estimation inversion problem (Moraal and Grizzle (1995b)) will be ill-posed (the solution does not exist, is not unique, or does not depend continuously on the data) or ill-conditioned (the unique solution is unacceptably sensitive to perturbations of the data), and particular consideration is required to achieve a robust estimator. Such modifications are generally known as regularization methods, see Tikhonov and Arsenin (1977). A common method, Tikhonov and Arsenin (1977), is to augment the cost function with a penalty on deviation from a priori information and makes the estimated solution degrade gracefully when $U_t$ is not $N$-exciting.

#### 1.3.1.2 Objective Function and Constraints

The topic of this section is to formulate the NMHE problem in terms of a nonlinear optimization problem that is convenient to solve using numerical opti-

mization. Defining $W_k = \text{col}(w_{k-N}..., w_{k-1}, w_k)$, and $V_k = \text{col}(v_{k-N}, ..., v_k)$ we introduce the following cost function similar to Rao et al (2003)

$$J'(x_{k-N}, ..., x_k, W_k, V_k) = \sum_{i=k-N}^{k} L(w_i, v_i) + Z_{k-N}(x_{k-N}) \tag{1.74}$$

where $L(w, v)$ is a stage cost typically of the least-squares type $L(w, v) = ||w||^2_M + ||v||^2_\Xi$ for some $M = M^T \succeq 0$ and $\Xi = \Xi^T \succeq 0$, there is a second term $Z$ that we will discuss shortly, and the minimization must be performed subject to the model constraints

$$x_{i+1} = F(x_i, u_i, w_i) \tag{1.75}$$
$$y_i = h(x_i, u_i) + v_i \tag{1.76}$$

and the additional constraints resulting from (1.60)-(1.62)

$$\text{col}(x_i, w_i, v_i) \in X \times W \times V, \quad i = k - N, ..., k \tag{1.77}$$
$$C(x_i, w_i, v_i, t_i) \leq 0, \quad i = k - N, ..., k \tag{1.78}$$
$$\sum_{i=k-N}^{k} c(x_i, w_i, v_i) \leq 0 \tag{1.79}$$

It is straightforward to eliminate the variables $v_i$ from this optimization problem, leading to

$$\begin{aligned} \Phi^*_{k-N} &= \min_{x_{k-N}, ...x_k, W_k} J(x_{k-N}, ..., x_k, W_k) \\ &= \sum_{i=k-N}^{k} L(w_i, y_i - h(x_i, u_i)) + Z_{k-N}(x_{k-N}) \end{aligned} \tag{1.80}$$

subject to

$$x_{i+1} = F(x_i, u_i, w_i), \quad i = k - N, ..., k \tag{1.81}$$
$$\text{col}(x_i, w_i, v_i) \in X \times W \times V, \quad i = k - N, ..., k \tag{1.82}$$
$$C(x_i, w_i, y_i - h(x_i, u_i), t_i) \leq 0, \quad i = k - N, ..., k \tag{1.83}$$
$$\sum_{i=k-N}^{k} c(x_i, w_i, y_i - h(x_i, u_i)) \leq 0 \tag{1.84}$$

By defining the solution function $\phi(i, U_k, x_{k-N})$ for $i \geq k - N$ using (1.81) recursively we can make further elimination of the nonlinear equality constraints (1.81) similar to the direct single shooting approach and re-define the cost function and constraints as follows:

$$\min_{x_{k-N},W_k} J(x_{k-N},W_k) = \sum_{i=k-N}^{k} L(w_i, y_i - h(\phi(i,U_k,x_{k-N}),u_i)) + Z_{k-N}(x_{k-N}) \tag{1.85}$$

subject to

$$\begin{aligned} \mathrm{col}(x_i,w_i,v_i) &\in X \times W \times V, \quad i = k-N,...,k \\ C(\phi(i,U_k,x_{k-N}),w_i,y_i - h(\phi(i,U_k,x_{k-N}),u_i),t_i) &\leq 0, \quad i = k-N,...,k \\ \sum_{i=k-N}^{k} c(\phi(i,U_k,x_{k-N}),w_i,y_i - h(\phi(i,U_k,x_{k-N}),u_i)) &\leq 0 \end{aligned} \tag{1.86}$$

The simple choice $Z(\cdot) = 0$ means that the state estimate is defined as the best least squares match with the data on the horizon. This means that no information from the data before the start of the horizon is used in the estimation, which is a clear weakness especially when the information content in the data is low due to lack of excitation, noise and other uncertainty. In other words, the estimation formulation contains no other mechanisms to introduce filtering of noise or regularization than to increase the horizon $N$, which also increases the computational complexity of the optimization problem and may still be insufficient.

In order to improve our ability to tune the NMHE and systematically introduce filtering of the state estimates, the term $Z(\cdot)$ in the formulation may be used as an arrival-cost estimate as discussed in e.g. Rao et al (2003) or in an ad hoc way to penalize deviation from an a priori estimate as in e.g Alessandri et al (2008); Sui and Johansen (2010); Alessandri et al (2003). Arrival cost estimation is discussed further in section 1.3.1.3, and a link between arrival cost estimation and the approach of Alessandri et al (2008) is illustrated in Poloni et al (2010).

We remark that the formulations make no particular assumptions on the uncertainty, and minimizes the impact of uncertainty on the estimates in a least-squares sense. Introduction of stochastic models can be envisioned and lead to better estimates in some cases, Lima and Rawlings (2010).

#### 1.3.1.3 Arrival-cost Estimates

The term $Z(\cdot)$ in the cost function $J$ defined in (1.80) may be used to make the finite (moving) window cost function $J$ approximate the full (still finite) window cost (Rao et al (2003))

$$J''(x_{k-N},...,x_k,W_k) = \sum_{i=0}^{k} L(w_i, y_i - h(x_i,u_i)) + \Gamma(x_0) \tag{1.87}$$

such that

$$Z_{k-N}(x_{k-N}) \approx \sum_{i=0}^{k-N-1} L(w_i, y_i - h(x_i, u_i)) + \Gamma(x_0) \tag{1.88}$$

where $\Gamma(x_0)$ is such that $\Gamma(x_0) = 0$ for the a priori most likely estimate of $x_0$, and $\Gamma(x) \succ 0$ for other values. The motivation for more closely approximating the full window cost (as opposed to a moving window cost) is to capture as much information as possible from time index $i = 0, 1, ...., k - N - 1$. Using arguments of dynamic programming, Rao et al (2003), an exact arrival cost completely captures the information up to time index $k - N - 1$. This would lead to more accurate estimates through improved filtering.

The effect of the arrival cost can be understood by comparing the moving horizon approach to Extended Kalman Filtering (EKF); Gelb (2002). In an EKF the information in past data is summarized in the covariance matrix estimate. Under assumptions that include linearity of the system and the noise and disturbances being Gaussian white noise with known covariances that are reflected in a quadratic cost function, it is known that the Kalman filter is an optimal filter, Gelb (2002), that provides states estimates with minimum variance. An EKF is an approximate sub-optimal filter that allows for nonlinearities and makes certain simplifying computations such neglecting higher order statistics and higher order (nonlinear) terms. In a similar manner, the NMHE with an arrival cost estimate captures the information of data until the start of the window in the arrival cost. Unfortunately, it is hard to find an explicit representation of the arrival cost for nonlinear systems, and practical methods attempts to approximate the arrival cost. The use of covariance matrix estimates from EKF and similar ideas is a useful way to define the arrival cost, Rao et al (2003):

$$Z_k(x) = (x - \hat{x}_k)^T \Pi_k^{-1} (x - \hat{x}_k) + \Phi_k^* \tag{1.89}$$

The matrix $\Pi_k$ is assumed to be non-singular such that its inverse is well defined, and obtained by solving the recursive Riccati-equation

$$\Pi_{k+1} = G_k Q_k G_k^T + A_k \Pi_k A_k^T - A_k \Pi_k C_k^T \left( R_k + C_k \Pi_k C_k^T \right)^{-1} C_k \Pi_k A_k^T$$

with some given positive definite matrix as initial condition $\Pi_0$. The matrices $A_k, G_k, C_k$ are defined as linearizations about the NMHE estimated trajectory:

$$A_k = \frac{\partial F(\hat{x}_k, u_k, \hat{w}_k)}{\partial x} \tag{1.90}$$

$$G_k = \frac{\partial F(\hat{x}_k, u_k, \hat{w}_k)}{\partial w} \tag{1.91}$$

$$C_k = \frac{\partial h(\hat{x}_k, u_k)}{\partial x} \tag{1.92}$$

and for simplicity we assume $Q_k$ and $R_k$ are defined through a quadratic cost function $L(w,v) = w^T Q_k^{-1} w + v^T R_k^{-1} v$. More generally, $Q_k$ and $R_k$ may be defined as Hessians of $L$ as in Rao et al (2003).

It is well known that alternative nonlinear Kalman Filters may perform better than the EKF in many situations. In the context of NMHE arrival cost estimation some useful methods are sample based filters (Ungarala (2009)), particle filtes (Lopez-Negrete et al (2009)), and Unscented Kalman Filtering (UKF) (Qu and Hahn (2009)).

#### 1.3.1.4 Combined State and Parameter Estimation

Many practical estimation problems are characterized by both states and parameters being unknown or uncertain. In Kalman filtering (Gelb (2002)) and observer design, a common approach to the joint state and parameter estimation problem is to augment the state space with constant parameters. Assuming a vector of constant parameters $\theta^*$ appears in the model equations:

$$\xi_{i+1} = F_m(\xi_i, u_i, \omega_i, \theta^*) \tag{1.93}$$

$$y_i = h_m(\xi_i, u_i, \theta^*) + v_i \tag{1.94}$$

with the new notation where $\xi_i$ is the state and $\omega_i$ is the disturbance. An *augmented* state space model assumes that the parameters are constant or slowly time-varying by the following model of the unknown parameter vector

$$\theta_{i+1} = \theta_i + \varrho_i \tag{1.95}$$

Combining (1.93)-(1.94) with (1.95) leads to

$$\begin{pmatrix} \xi_{i+1} \\ \theta_{i+1} \end{pmatrix} = \begin{pmatrix} F_m(\xi_i, u_i, \omega_i, \theta_i) \\ \theta_i + \varrho_i \end{pmatrix} \tag{1.96}$$

$$y_i = h_m(z_i, u_i, \theta_i) + v_i \tag{1.97}$$

With the augmented state $x = \mathrm{col}(\xi, \theta)$ and augmented disturbance vector $w = \mathrm{col}(\omega, \varrho)$ we observe that these equations are in the assumed form (1.75)-(1.76) such that the NMHE algorithm formulation can be applied without any modifications.

It is common to encounter combined state and parameter estimation problems where convergence conditions of uniform observability or persistence of excitation are not fulfilled, Moraal and Grizzle (1995a); Sui and Johansen (2010). In such cases various mechanisms of regularization should be implemented to get graceful degradation of the estimation when insufficient information is available to determine the estimates. The use of a term in the cost function that preserves the history and makes the observer degrade to an open-loop observer is one such mechanism, that can be combined with more

advanced monitoring of the Hessian matrix of the cost function to detect and resolve lack of excitation, Sui and Johansen (2010).

## 1.4 Numerical Optimization

For simplicity of notation, we assume in this section that the NMHE or NMPC problem is formulated as a general nonlinear programming problem at each time instant

$$\min_z V(z) \text{ subject to } G(z) \leq 0,\ H(z) = 0 \tag{1.98}$$

where $z$ is a vector with the unknown decision variables. In practice, as implemented in most numerical solver software, it will be important to exploit structural properties of the constraints and objective functions such that further separation of the functions $G$ and $H$ into simple bounds ($z_{min} \leq z \leq z_{max}$), linear constraints and "truly" nonlinear constraints is usually made for efficient implementation. For simplicity of presentation, we does not make such separation here.

### *1.4.1 Problem Structure*

The choice of numerical optimization solver strategy will have significant impact on both the need for computational resources and the quality of the solution in NMPC and NMHE. In this context, computational resources usually means the CPU time required for the solution to converge to meet the tolerance requirements, while quality of solution is related to lack of convergence or high sensitivity to initial guesses.

There are several features of NMPC and NMHE problems that are relevant to consider

- Formulation of the numerical optimal control or estimation problem, e.g. sequential or simultaneous approaches. The sequential approach leads to a smaller, denser problem with a computationally complex cost function usually without nonlinear equality constraints, while the simultaneous approach leads to a larger, more structured, sparse problem with nonlinear equality constrains and relatively simple cost and constraint functions to evaluate.
- NMPC and NMHE solves a sequence of numerical optimal control or estimation problems, where the parameters of the problem are usually subject to fairly small changes from one run to the next. There is usually benefits of warm starting the next optimization run using the solution and other internal data from the previous run as initial guesses, data or conditions.

- Since the optimization will be repeated at the next sample, and the optimization problem is formulated using uncertain data, it may not always be essential that the solver has converged (or equivalently that the tolerances may not need to be very strict) due to the forgiving effect of feedback. However, a feasible solution is generally required at each run in order to operate the control and monitoring systems. This means that problems tend to be re-formulated using slack variables with some prioritization of constraints that can be relaxed, and that is it generally desirable to start the next optimization run with a feasible initial guess generated from the previous run such that even with a limited number of iterations one can guarantee feasibility.
- Safety and reliability are essential features of most control and monitoring systems, which means that post-optimal analysis and checks on the quality of the solution must usually be implemented. Issues such as non-convexity and non-smoothness of models and constraints are essential to understand and take into account.

Although all nonlinear MPC and MHE problems have certain features in common, they may also differ considerably with respect to size, models, cost functions and constraints. This means that there will not be a single numerical method that will be the best, in general. Below, we briefly outline some commonly used numerical methods with emphasis on sequential quadratic programming and interior point methods. We point out that there exist a wide range of alternative methods that may perform better in certain types of problems, like derivative-free methods (e.g. Conn et al (2009)) that may be better suited if the computation of gradients is expensive or not possible to achieve accurately.

### *1.4.2 Nonlinear Programming*

Newton's method for iterative solution of nonlinear algebraic equations is the backbone of most numerical optimization methods. For a nonlinear vector equation $f(z) = 0$, Newton's method starts with an initial guess vector $z^0$ and generates a sequence of guesses $z^k$ indexed by the integer $k = 1, 2, 3, ....$ according to the following formula that results from linearization using Taylor's theorem and truncation:

$$f(z^k) + \nabla_z^T f(z^k)(z^{k+1} - z^k) = 0 \tag{1.99}$$

Eq. (1.99) defines a set of linear algebraic equations that can be solved for $z^{k+1}$ using numerical linear algebra, which is the workhorse at the core of nonlinear programming and is the main contribution to computational complexity in addition to the computation of the function $f$ and its gradient (Jacobian matrix) $\nabla_z f$. As Newton's method is based on linearization, it has

only local convergence, but with a quadratic convergence rate, Nocedal and Wright (1999).

Newton's method is used in nonlinear programming to solve nonlinear algebraic equations closely related to the first order optimality conditions of (1.98), known as the Karush-Kuhn-Tucker (KKT) conditions Nocedal and Wright (1999)

$$\nabla_z L(z^*, \lambda^*, \mu^*) = 0 \tag{1.100}$$
$$H(z^*) = 0 \tag{1.101}$$
$$G(z^*) \leq 0 \tag{1.102}$$
$$\mu^* \geq 0 \tag{1.103}$$
$$G_i(z^*)\mu_i^* = 0, \ i = 1, ...., n_G \tag{1.104}$$

where $n_G$ is the number of inequality constraints and the Lagangian function is defined as

$$L(z, \lambda, \mu) = V(z) + \lambda^T H(z) + \mu^T G(z) \tag{1.105}$$

Obviously, the KKT conditions also involves inequalities which means that Newton's method cannot be applied directly. The different nonlinear programming methods differ conceptually in the way the KKT conditions, being mixed equations and inequalities, are used to formulate a sequence of nonlinear equations. The different nonlinear programming methods also differ with respect to approximations used for the gradient $\nabla_z f$ of the resulting set of equations. Since the evaluation of (1.100) already requires gradient computations (for the Jacobian matrix of the Lagrangian $\nabla_z L$) in the formulation of the equations to be solved, the computation of $\nabla_z f$ generally requires the expensive computation or approximation of the matrix $\nabla_z^2 L$, known as the Hessian matrix of the Lagrangian.

#### 1.4.2.1 Sequential Quadratic Programming (SQP)

SQP methods linearize the KKT conditions (1.100)-(1.104) at the current iterate $z^k$, leading to a set of linear conditions that can be interpreted as the KKT conditions of the following quadratic program (QP), Nocedal and Wright (1999):

$$\min_z V_{QP}^k(z) \tag{1.106}$$

subject to

$$H(z^k) + \nabla_z^T H(z^k)(z - z^k) = 0 \tag{1.107}$$
$$G(z^k) + \nabla_z^T G(z^k)(z - z^k) \leq 0 \tag{1.108}$$

with the cost function

$$V_{QP}^k(z) = \nabla_z^T V(z^k)(z - z^k) + \frac{1}{2}(z - z^k)^T \nabla_z^2 L(z^k, \lambda^k, \mu^k)(z - z^k) \tag{1.109}$$

This QP interpretation is highly useful since it provides a practical way to deal with the fact that the KKT conditions include inequalities, which are not straightforward to solve using Newton's method directly. The vast knowledge and numerical methods of solving QP problems, typically using active set methods, Nocedal and Wright (1999); Gill et al (1981), is exploited at this point. Active set methods replace inequality constraints with equality constraints based on an active set assumption that is improved iteratively as the method converges towards an optimal solution.

However, there are three major challenges remaining:

- The first key challenge is related to the Hessian matrix $\nabla_z^2 L(\cdot)$. Problems arise if this matrix is not positive definite such that the QP is not convex and a global optimum may not exist or is not unique. In the context of NMPC or NMHE, problems will also arise if the computational complexity of computing the Hessian is beyond the CPU resources available. Approximations such as quasi-Newton and Gauss-Newton methods are commonly used to approximate the Hessian from the Jacobian, see below, in a positive definite form.
- The second key challenge is related related to the accuracy of the underlying linearizations (or equivalently, the local quadratic approximations of the QP to the NLP). In order to have control over this issue, it is common to solve the QP to generate a search direction only, and then generate the next iterate $z^{k+1}$ not as the minimum of the QP defined above, but through a search procedure along this direction. Common search procedures are line search and trust region methods, as outlined below.
- The third key challenge is related to feasibility. To ensure convergence it is common to use a merit function to control the step size length in both line search and trust region methods. The merit function adds a penalty on constraint violations to the original cost function to ensure that the next iterate moves towards a combined objective of reducing the cost function and being feasible.

**Quasi-Newton methods** approximate the Hessian of the Lagrangian by an update formula that only requires computation of the Jacobian. Common methods, such as the BFGS update, Nocedal and Wright (1999), leads to significant computational reduction and ensures that the Hessian approximation is positive definite. The price to pay is that the convergence rate may no longer be quadratic, but typically only super-linear, Nocedal and Wright (1999).

**Gauss-Newton methods** are particularly useful for least-squares type of problems, like NMHE and certain NMPC formulations, where the cost function is the squared norm of some nonlinear functions since a reliable

estimate of the Hessian can be computed directly from the Jacobian as the product of the Jacobian and its transpose, Nocedal and Wright (1999).

**Line search methods** are designed to account for the fact that the QP is only a locally valid approximation. As the name indicates, one performs a one-dimensional search in the descent direction computed by the QP (solution) to ensure that sufficient descent of the actual merit function is achieved; Nocedal and Wright (1999).

**Trust region methods** define a maximum step length for the next iterate based on a trust region, where the linearization is sufficiently accurate. This aims to ensure that the next iterate is well defined and accurate, and the size of the trust region is adapted to ensure that the merit function reduction predicted by the QP is sufficiently close to the actual merit function reduction, Conn et al (2000); Wright and Tenny (2004).

#### 1.4.2.2 Interior Point Methods (IP)

Interior point methods deal with the inequality constraints of the KKT conditions in a fundamentally different way than SQP methods. The KKT conditions concerning the inequality constraints, in particular (1.104), is replaced by a smooth approximation (Wright (1997); Diehl et al (2009)):

$$G_i(z^*)\mu_i^* = \tau,\ i = 1, ...., n_G \tag{1.110}$$

Solving the resulting set of algebraic nonlinear equations with Newton's methods is equivalent to a solution of the following appoximate problem, where the inequality constraints are handled by a $\log(\cdot)$ barrier function:

$$\min_z \left( V(z) - \tau \sum_{i=1}^{n_G} \log\left(-G_i(z)\right) \right) \text{ subject to } H(z) = 0 \tag{1.111}$$

The parameter $\tau > 0$ parameterizes a central path in the interior of the feasible region towards the optimum as $\tau \to 0$, which motivates the name of IP methods. Once the solution for a given $\tau > 0$ is found, the parameter $\tau$ can be reduced by some factor in the next Newton iteration. The practical implementation of an IP method will typically use Newton's method to compute a search direction. Challenges related to the computation of the Hessian matrix and limited validity of the linearization of the Newton method, remain similar to SQP, and the ideas of quasi-Newton methods, merit functions, line search and trust regions are relevant and useful also for IP methods.

#### 1.4.2.3 Linear Algebra

At heart of both the QP sub-problems of SQP and the Newton-step of IP methods are the solution of a set of linear algebraic equations. Efficiency of the numerical optimization solver heavily depends on the efficiency of solving this problem, since it will be repeated many times towards the solution of the NLP at each sampling instant of an NMPC or NMHE. Exploiting structural properties is essential.

Depending on the solution strategy and properties of the problem, such structural properties are often related to positive definiteness of the Hessian (approximation), sparseness and block-diagonal structure of the linear systems of equations, and what information from the previous optimization run can be used to initialize the next run. Using factorization methods one may eliminate algebraic variables and operate in reduced spaces to save computations. Being able to efficiently maintain and update factorized matrices between the various iterations is usually essential to implement this. Although this is essential in any practical implementation of NMHE and NMPC, it is a fairly complex bag of tricks and tools that we consider outside the scope of this introduction. Instead, we refer to excellent and comprehensive treatments in Nocedal and Wright (1999); Diehl et al (2009); Gill et al (1997, 1981) and the references therein.

### *1.4.3 Warm Start*

The NLP problem at one sampling instant is usually closely related to the NLP problem at the previous sampling instant in NMPC and NMHE problem, since the sampling interval is usually short compared to the dynamics of the plant and the controller. Assuming the reference signals and other input to the controller changes slowly, this means that the solution in terms of past state trajectories (for MHE problems) or future input and state trajectories (for MPC problems) can be time shifted one sampling period and still provide a reasonably accurate solution to the next NLP. Assuming no uncertainty in MPC problems, this is a perfectly valid assumption and is commonly used to guarantee feasibility at the next step in stability arguments, e.g Scokaert et al (1999); Mayne et al (2000). Even without time-shifting, the previous solution itself also provides a good initialization for warm start purposes in NMPC, Boch et al (1999); Diehl et al (2004).

Unlike SQP methods, IP methods can usually not make effective use of initial guesses of the solution due to the reformulation of the KKT conditions that follows the parameterized center path controlled by the parameter $\tau > 0$ that is sequentially reduced towards zero. This does not necessarily imply that IP methods are less suited for NMPC and NMHE problems, in particular for large scale problems where IP methods have advantages that may compensate

for this shortcoming. Modified IP methods that can efficiently incorporate warm start is a current research topic, Gondzio and Grothey (2008); Shahzad et al (2010).

Warm start is potentially most efficient when including data beyond just the solution point, but also consider the internal data of the optimization algorithm such as initial estimates of the Hessian approximation (in case exact Hessians are not computed), or initial estimates of factorizations of the Hessian (approximation), initial estimates of optimal active sets, and other data. This is in particular a challenge when the dimensions and structure of these internal data will change from one sample to the next. This may for example be the case in the simultaneous formulations (in particular direct collocation) of numerical optimal control (see section 1.2.2), since the discretization may be changed from one sample to the next, in general. One must also have in mind that simultaneous formulations require that both state and control trajectories are initialized, while sequential formulations only require the control trajectory initialization. What is most beneficial will depend on the accuracy of the available information for initialization, amongst other things. We refer to Diehl et al (2009); Houska et al (2010) and the references therein for a deeper treatment of this topic.

### *1.4.4 Computation of Jacobians and Hessians*

The computation of the Jacobians of the cost and constraint functions is often the main computational cost of numerical optimization methods, and even fairly small inaccuracies in the calculation of the Jacobians due to may lead to severe convergence problems.

Simultaneous approaches offer advantages over sequential approaches with respect to Jacobian computations:

- The prediction horizon is broken up into several intervals where ODE solutions are computed from given initial conditions. Since these intervals will be shorter than the single interval of a single shooting approach, numerical errors due to the ODE solver tend to accumulate less.
- Implicit ODE solvers, which generally have more stable numerical properties than explicit solvers, can in general be used in simultaneous approach.
- Simultaneous approaches are characterized by simpler cost and constraint functions, where automatic differentiation is more easily exploited to avoid numerical Jacobian computation errors, see section 1.4.4.2.

The numerical challenges are in particular important to consider for plants that are unstable or marginally stable. Like in linear MPC, there may be advantages of pre-stabilizing an open-loop unstable plant model with a feedback compensator before used in NMPC or NMHE, Cannon and Kouvaritakis (2005); Sui et al (2010).

#### 1.4.4.1 Finite Difference

The finite difference method approximates the $(i,j)$-th element of the Jacobian of a vector function $f(z)$ as

$$(\nabla_z f(z))_{i,j} \approx \frac{f_i(z_j+\delta) - f_i(z_j)}{\delta} \tag{1.112}$$

for some small $\delta > 0$. If $\delta$ is too large there will be inaccuracies due to the nonlinearity of $f_i$, since the method computes the average slope between two points. If the two points are not infinitely close and the function is not linear, there will be a "nonlinearity error". If $\delta$ is too small, any finite numerical error $\varepsilon_1$ in the computation of $f_i(z_j+\delta)$ and $\varepsilon_2$ in the computation of $f_i(z_j)$ will lead to an error $\epsilon = (\varepsilon_1 - \varepsilon_2)/\delta$ in the computation of the derivative. Obviously, this error goes to infinity when $\delta \to 0$, so a tradeoff between these errors must be made. It should be noticed that the finite difference approximation error $\epsilon$ depends on the difference between the errors in the two point-wise evaluations of $f_i$. This means that systematic errors (i.e. the same error in both $\varepsilon_1$ and $\varepsilon_2$) will have a much smaller effect than a random error of the same magnitude. Practical experience shows that the use of variable-step (adaptive) ODE solvers tend to give a small random numerical error, while the use of fixed-step ODE solvers tend to give a larger systematic error, but even smaller random error. For the reasons described above, one may find that a fixed-step ODE solver leads to considerably smaller error in finite difference Jacobian computations and performs better with less convergence problems in many numerical methods for NMPC and NMHE.

It is also worthwhile to remind the reader that scaling of all variables involved in the optimization problem to the same order of magnitude is in many cases a pre-requisite for numerical nonlinear optimization methods to work satisfactorily. This is evident in the context of finite difference Jacobian computations, but also relevant for other numeric computations.

As a final remark, it is possible to exploit square-root factorizations (like Cholesky factorization) for improved numerical accuracy and computational complexity in finite difference computations, Schei (1997).

#### 1.4.4.2 Symbolic and Automatic Differentiation

The most accurate result and computationally most efficient approach is to calculate gradients by symbolically differentiating the cost and constraint functions. Doing this by hand, or even using symbolic computations in Matlab, Maple or Mathematica, may easily become intractable for NMPC and NMHE problems that may contain a large number of variables, equations and inequalities. A more convenient solution is to rely on so-called *automatic differentiation* software (Griewank and Walther (2008)) that achieved this objective either by overlaying operators in object oriented languages such

as C++ (Griewank et al (1996)), or automatically generates source code for gradient functions based on source code of the original function, Bischof et al (1996).

**Acknowledgements** The work is supported by a grant No. NIL-I-007-d from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism and the Norwegian Financial Mechanism.

## References

Abraham R, Marsden JE, Ratiu T (1983) Manifolds, Tensor Analysis, and Applications. Springer–Verlag New York

Adetola V, Guay M (2003) Nonlinear output feedback receding horizon control. In: Proc. ACC, Denver

Åksesson BM, Toivonen HT (2006) A neural network model predictive controller. Journal of Process Control 16:937–946

Alamir M (1999) Optimization based non–linear observers revisited. Int J Control 72:1204–1217

Alamir M (2007) Nonlinear moving horizon observers: Theory and real-time implementation. In: Besancon G (ed) Nonlinear Observers and Applications, LNCIS 363, Springer, pp 139–179

Alamir M, Bornard G (1995) Stability of a truncated infinite constrained receding horizon scheme: The general discrete nonlinear case. Automatica 31:1353–1356

Alessandri A, Baglietto M, Parisini T, Zoppoli R (1999) A neural state estimator with bounded errors for nonlinear systems. IEEE Transactions on Automatic Control 44:2028 – 2042

Alessandri A, Baglietto M, Battistelli G (2003) Receding–horizon estimation for discrete–time linear systems. IEEE Transactions Automatic Control 48:473–478

Alessandri A, Baglietto M, Battistelli G (2008) Moving–horizon state estimation for nonlinear discrete–time systems: New stability results and approximation schemes. Automatica 44:1753–1765

Allgöwer F, Badgwell TA, Qin JS, Rawlings JB, Wright SJ (1999) Nonlinear predictive control and moving horizon estimation – An introductory overview. In: Frank PM (ed) Advances in Control, Highlights of ECC-99, Springer, pp 391–449

Angeli D, Casavola A, Mosca E (2000) Constrained predictive control of nonlinear plants via polytopic linear system embedding. Int J Robust Nonlinear Control 10:1091 – 1103

Arellano-Garcia H, Wendt M, Barz T, Wozny G (2007) Closed-loop stochastic dynamic optimization under probabilistic output constraints. In: Findeisen R, Allgöwer F, Biegler LT (eds) Assessment and future directions of nonlinear model predictive control, LNCIS, vol. 358, Springer-Verlag, pp 305–315

Athans M, Falb PL (1966) Optimal Control. An Introduction to the Theory and Its Applications. McGraw–Hill Ltd.

Balchen JG, Ljungquist D, Strand S (1992) State–space predictive control. Chemical Engineering Science 47:787–807

Bellman R (1957) Dynamic Programming. Princeton University Press, New Jersey

Bemporad A, Filippi C (2003) Suboptimal explicit RHC via approximate multiparametric quadratic programming. J Optimization Theory and Applications 117:9–38

Bemporad A, Morari M (1999) Control of systems integrating logic, dynamics, and constraints. Automatica 35:407–427

Bemporad A, Borrelli F, Morari M (2000) Optimal controllers for hybrid systems: Stability and piecewise linear explicit form. In: Proc. Conference on Decision and Control, Sydney

Bemporad A, Morari M, Dua V, Pistikopoulos EN (2002) The explicit linear quadratic regulator for constrained systems. Automatica 38:3–20

Betts JT (2001) Practical methods for optimal control using nonlinear programming. SIAM, Philadelphia

Biegler L (2000) Efficient solution of dynamic optimization and NMPC problems. In: Allgöwer F, Zheng A (eds) Nonlinear Predictive Control, Birkhäuser, pp 219 – 244

Biegler LT (1984) Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. Computers and Chemical Engineering 8 pp 243–248

Bischof C, Carle A, Kadhemi P, Mauer A (1996) ADIFOR2.0: Automatic differentiation of Fortran 77 programs. IEEE Computational Science and Engineering 3:18–32

Biyik E, Arcak M (2006) A hybrid redesign of Newton observers in the absence of an exact discrete time model. Systems and Control Letters 55:429–436

Bølviken E, Acklam PJ, Christophersen N, Størdal JM (2001) Monte Carlo filters for non–linear state estimation. Automatica 37:177–183

Boch HG, Diehl M, Leineweber DB, Schlöder JP (1999) Efficient direct multiple shooting in nonlinear model predictive control. In: Keil F, Mackens W, Voss H, Werther J (eds) Scientific computing in chemical engineering, vol 2, Springer, Berlin

Bock HG, Plitt KJ (1984) A multiple shooting algorithm for direct solution of optimal control problems. In: Proceedings 9th IFAC World Congress, Budapest, Pergamon Press, Oxford, pp 243–247

Bock HG, Diehl M, Leineweber DB, Schlöder JP (1999) Efficient direct multiple shooting in nonlinear model predictive control. In: Keil F, Mackens W, Voß H, Werther J (eds) Scientific Computing in Chemical Engineering II, vol 2, Springer, Berlin, pp 218–227

Borrelli F, Morari M (2007) Offset free model predictive control. In: Proc. IEEE Conference on Decision and Control, pp 1245–1250

Boyd S, Ghaoui LE, Feron E, Balahrishnan V (1994) Linear Matrix Inequalities in System and Control Theory. SIAM, Philadelphia

Cannon M, Kouvaritakis B (2005) Optimizing prediction dynamics for robust mpc. IEEE Trans Automatic Control 50:1892–1897

Cannon M, Ng D, Kouvaritakis B (2009) Successive linearization NMPC for a class of stochastic nonlinear systems. In: Magni L, Raimondo DM, Allgöwer F (eds) Nonlinear Model Predictive Control: Towards New Challenging Applications, LNCIS, vol. 384, Berlin/Heidelberg: Springer-Verlag, pp 249–262

Casavola A, Famularo D, Franze G (2003) Predictive control of constrained nonlinear systems via LPV linear embeddings. Int J Robust Nonlinear Control 13:281 – 294

Cervantes A, Biegler L (1998) Large–scale DAE optimization using a simultaneous nlp formulation. AIChE Journal 44:1038–1050

Chen CC, Shaw L (1982) On receding horizon feedback control. Automatica 18:349–352

Chen H, Allgöwer F (1998) A quasi–infinite horizon nonlinear model predictive control scheme with guaranteed stability. Automatica 34:1205–1217

Conn A, Scheinberg K, Vicente LN (2009) Introduction to Derivative-Free Optimization. SIAM

Conn AR, Gould NIM, Toint PL (2000) Trust region methods. SIAM

Deuflhard P (1974) A modified Newton method for the solution of ill–conditioned systems of nonlinear equations with application to multiple shooting. Numer Math 22:289–315

Diehl M, Bock HG, Schlöder JP, Findeisen R, Nagy Z, Allgöwer F (2002) Real–time optimization and nonlinear model predictive control of processes governed by differential–algebraic equations. J Process Control 12:577–585

Diehl M, Magni L, de Nicolao G (2004) Online NMPC of unstable periodic systems using approximate infinite horizon closed loop costing. Annual Reviews in Control 28:37–45

Diehl M, Ferreau HJ, Haverbeke N (2009) Efficient numerical methods for nonlinear MPC and moving horizon estimation. In: et al LM (ed) Nonlinear Model Predictive Control, LNCIS 384, Springer, pp 391–417

Dunbar WB, Murray RM (2006) Distributed receding horizon control for multi-vehicle formation stabilization. Automatica 42:549–558

Feng L, Gutvik CR, Johansen TA, Sui D (2009) Barrier function nonlinear optimization for optimal decompression of divers. In: IEEE Conf. Decision and Control, Shanghai

Fiacco AV (1983) Introduction to sensitivity and stability analysis in nonlinear programming. Orlando, Fl: Academic Press

Findeisen R, Imsland L, Allgöwer F, Foss BA (2003a) Output feedback stabilization for constrained systems with nonlinear model predictive control. Int J Robust and Nonlinear Control 13:211–227

Findeisen R, Imsland L, Allgöwer F, Foss BA (2003b) State and output feedback nonlinear model predictive control: An overview. European J Control 9:179–195

Foss BA, Schei TS (2007) Putting nonlinear model predictive control into use. In: Assessment and Future Directions Nonlinear Model Predictive Control, LNCIS 358, Springer Verlag, pp 407 – 417

Foss BA, Johansen TA, Sørensen AV (1995) Nonlinear predictive control using local models – applied to a batch fermentation process. Control Engineering Practice 3:389–396

Gelb A (2002) Applied Optimal Estimation, 17th edn. MIT Press

Gill P, Murray W, Wright M (1981) Practical optimization. Academic Press, Inc.

Gill P, Barclay A, Rosen JB (1997) SQP methods and their application to numerical optimal control. Tech. Rep. NA 97–3, Department of Mathematics, University of California, San Diego

Glad ST (1983) Observability and nonlinear dead beat observers. In: IEEE Conf. Decision and Control, San Antonio, pp 800–802

Gondzio J, Grothey A (2008) A new unblocking technique to warmstart interior point methods based on sensitivity analysis. SIAM J Optimization 19:1184âĂŞ1210

Goodwin GC, De Dona JA, Seron MM, Zhuo XW (2005) Lagrangian duality between constrained estimation and control. Automatica 41:935–944

Goodwin GC, Østergaard J, Quevedo DE, Feuer A (2009) A vector quantization approach to scenario generation for stochastic NMPC. In: Magni L, Raimondo DM, Allgöwer F (eds) Nonlinear Model Predictive Control: Towards New Challenging Applications, LNCIS, vol. 384, Berlin/Heidelberg: Springer-Verlag, pp 235–248

Gravdahl JT, Egeland O (1997) Compressor surge control using a close–coupled valve and backstepping. In: Proc. American Control Conference, Albuquerque, NM., vol 2, pp 982 –986

Greitzer EM (1976) Surge and rotating stall in axial flow compressors, part i: Theoretical compression system model. J Engineering for Power 98:190–198

Griewank A, Walther A (2008) Evaluating Derivatives, second edition. SIAM

Griewank A, Juedes D, Utke J (1996) ADOL–C, A package for the automatic differentiation of algorithms written in C/C++. ACM Trans Mathematical Software 22:131–167

Grimm G, Messina MJ, Tuna SE, Teel AR (2004) Examples when nonlinear model predictive control is nonrobust. Automatica 40:1729–1738

Haseltine EL, Rawlings JB (2005) Critical evaluation of extended Kalman filtering and moving–horizon estimation. Ind Eng Chem Res 44:2451–2460

Heemels WPMH, Schutter BD, Bemporad A (2001) Equivalence of hybrid dynamical models. Automatica 37:1085 – 1091

Hicks GA, Ray WH (1971) Approximation methods for optimal control systems. Can J Chem Engng 49:522–528

Houska B, Ferreau HJ, Diehl M (2010) ACADO toolkit – an open–source framework for automatic control and dynamic optimization. Optimal Control Applications and Methods

Isidori A (1989) Nonlinear Control Systems, 2nd Ed. Springer Verlag, Berlin

Jadbabaie A, Yu J, Hauser J (2001) Unconstrained receding–horizon control of nonlinear systems. IEEE Trans Automatic Control 46:776–783

Jang SS, Joseph B, Mukai H (1986) Comparison of two appraoches to on-linear parameter and state estimation of nonlinear systems. Ind Chem Proc Des Dev 25:809–814

Jazwinski AH (1968) Limited memory optimal filtering. IEEE Trans Automatic Control 13:558–563

Johansen TA (2004) Approximate explicit receding horizon control of constrained nonlinear systems. Automatica 40:293–300

Kandepu R, Foss B, Imsland L (2008) Applying the unscented Kalman filter for nonlinear state estimation. J Process Control 18:753–768

Kantas N, Maciejowski JM, Lecchini-Visintini A (2009) Sequential Monte Carlo for model predictive control. In: Magni L, Raimondo DM, Allgöwer F (eds) Nonlinear Model Predictive Control: Towards New Challenging Applications, LNCIS, vol. 384, Berlin/Heidelberg: Springer-Verlag, pp 263–274

Keerthi SS, Gilbert EG (1988) Optimal infinite horizon feedback laws for a general class of constrained discrete–time systems: Stability and moving horizon approximations. J Optimization Theory and Applications 57:265–293

Kerrigan E, Maciejowski JM (2000) Invariant sets for constrained nonlinear discrete–time systems with application to feasibility in model predictive control. In: Proc. IEEE Conf. Decision and Control, Sydney

Keviczky T, Borrelli F, Balas GJ (2006) Decentralized receding horizon control for large scale dynamically decoupled systems. Automatica 42:2105 – 2115

Kim I, Liebman M, Edgar T (1991) A sequential error-in-variables method for nonlinear dynamic systems. Comp Chem Engr 15:663–670

Kojima M (1980) Strongly stable stationary solutions in nonlinear programs. In: Robinson SM (ed) Analysis and Computation of Fixed Points, Academic Press, New York, pp 93–138

Kolås S, Foss B, Schei TS (2008) State estimation IS the real challenge in NMPC. In: Int. Workshop on Assessment and Future Directions of NMPC, Pavia, Italy

Kraft D (1985) On converting optimal control problems into nonlinear programming problems. In: Schittkowski K (ed) Computational Mathematical Programming, vol F15, NATO ASI Series, Springer–Verlag, pp 261–280

Krstic M, Kanellakopoulos I, Kokotovic P (1995) Nonlinear and Adaptive Control Design. Wiley and Sons

Lazar M, Heemels W, Bemporad A, Weiland S (2007) Discrete-time non-smooth nonlinear mpc: Stability and robustness. In: Findeisen R, Allgöwer F, Biegler LT (eds) Assessment and Future Directions of Nonlinear Model Predictive Control, Lecture Notes in Control and Information Sciences, Vol. 358, Springer Berlin / Heidelberg, pp 93 – 103

Lazar M, Heemels WPMH, Roset BJP, Nijmeijer H, van den Bosch PPJ (2008) Input-to-state stabilizing sub-optimal NMPC with an application to DC-DC converters. International Journal of Robust and Nonlinear Control 18:890 – 904

Leineweber DB, Bauer I, Boch HG, Schlöder JP (2003) An efficient multiple shooting based reduced SQP strategy for large–scale dynamic process optimization. part i: Theoretical aspects. Comp Chem Eng 27:157–166

Li WC, Biegler LT (1989) Multistep Newton-type control strategies for constrained nonlinear processes. Chem Eng Res Des 67:562–577

Li WC, Biegler LT (1990) Newton-type controllers for constrained nonlinear processes with uncertainty. Ind Engr Chem Res 29:1647–1657

Lima FV, Rawlings JB (2010) Nonlinear stochastic modeling to improve state estimation in process monitoring and control. AIChE Journal

Limon D, Alamo T, Camacho EF (2002) Input–to–state stable mpc for constrained discrete–time nonlinear systems with bounded additive uncertainties. In: Proc. IEEE Conference on Decision and Control, Las Vegas, NV, pp 4619–4624

Limon D, Alamo T, Salas F, Camacho EF (2006) Input–to–state stability of min–max MPC controllers for nonlinear systems with bounded uncertainties. Automatica 42:797–803

Lopez-Negrete R, Patwardhan SC, Biegler LT (2009) Approximation of arrival cost in moving horizon estimation using a constrained particle filter. In: 10th Int. Symposium Process Systems Engineering, pp 1299–1304

Magni L, Scattolini R (2006) Stabilizing decentralized model predictive control of nonlinear systems. Automatica 42:1231–1236

Magni L, Scattolini R (2007) Robustness and robust design of MPC for nonlinear discrete–time systems. In: Findeisen R, Allgöwer F, Biegler L (eds) Assessment and Future Directions of Nonlinear Model Predictive Control. Lecture Notes in Control and Information Sciences, vol 358, Springer–Verlag, pp 239–254

Magni L, Nicolao GD, Scattolini R (2001a) Output feedback and tracking of nonlinear systems with model predictive control. Automatica 37:1601âĂŞ1607

Magni L, Nicolao GD, Scattolini R (2001b) A stabilizing model-based predicitve control algorithm for nonlinear systems. Automatica 37:1351âĂŞ1362

Magni L, De Nicolao G, Scattolini R, Allgöwer F (2003) Robust model predictive control for nonlinear discrete–time systems. International Journal of Robust and Nonlinear Control 13:229–246

Mangasarian OL, Rosen JB (1964) Inequalities for stochastic nonlinear programming problems. Operations Research 12:143–154

Marino R, Tomei P (1995) Nonlinear Control Design: Geometric, Adaptive and Robust. Prentice Hall, UK

Mayne DQ, Michalska H (1990) Receding horizon control of nonlinear systems. IEEE Trans Automatic Control 35:814–824

Mayne DQ, Rawlings JB, Rao CV, Scokaert POM (2000) Constrained model predictive control: Stability and optimality. Automatica 36:789–814

Messina MJ, Tuna SE, Teel AR (2005) Discrete-time certainty equivalence output feedback: allowing discontinuous control laws including those from model predictive control. Automatica 41:617–628

Michalska H, Mayne DQ (1993) Robust receding horizon control of constrained nonlinear systems. IEEE Trans Automatic Control 38:1623–1633

Michalska H, Mayne DQ (1995) Moving horizon observers and observer–based control. IEEE Transactions Automatic Control 40:995–1006

Moraal PE, Grizzle JW (1995a) Asymptotic observers for detectable and poorly observable systems. In: IEEE Conf. Decision and Control, New Orleans, pp 109–114

Moraal PE, Grizzle JW (1995b) Observer design for nonlinear systems with discrete–time measurement. IEEE Transactions Automatic Control 40:395–404

Morari M, Lee J (1999) Model predicitve control: Past, present and future. Comp and Chem Eng 23:667–682

Muske KR, Badgwell TA (2002) Disturbance modeling for offset-free linear model predictive control. J Process Control 12:617 – 632

Nagy Z, Braatz RD (2003) Robust nonlinear model predictive control of batch processes. AIChE Journal 49:1776–1786

Nagy Z, Mahn B, Franke R, Allgöwer F (2007) Real-time implementation of nonlinear model predictive control of batch processes in an industrial framework. In: Findeisen R, Allgöwer F, Biegler LT (eds) Assessment and Future Directions of Nonlinear Model Predictive Control, Lecture Notes in Control and Information Sciences Vol. 358, Springer Berlin / Heidelberg, pp 465 – 472

Nicolao GD, Magni L, Scattolini R (2000) Stability and robustness of nonlinear receding horizon control. In: Allgöwer F, Zheng A (eds) Nonlinear Predictive Control, Birkhäuser, pp 3–23

Nijmeijer H, van der Schaft AJ (1990) Nonlinear Dynamical Control Systems. Springer–Verlag, New York

Nørgaard M, Ravn O, Poulsen NK, Hansen LK (2000) Neural Networks for Modelling and Control of Dynamic Systems. Springer, London

Nocedal J, Wright SJ (1999) Numerical Optimization. Springer–Verlag, New York

Ohtsuka T (2004) A continuation/GMRES method for fast computation of nonlinear receding horizon control. Automatica 40:563–574

Pannocchia G, Bemporad A (2007) Combined design of disturbance model and observer for offset-free model predictive control. IEEE Trans Automatic Control 52:1048–1053

Pannocchia G, Rawlings JB (2003) Disturbance models for offset-free model-predictive control. AIChE J 49:426 – 437

Poloni T, Rohal-Ilkiv B, Johansen TA (2010) Damped one–mode vibration model state and parameter estimation via pre–filtered moving horizon observer. In: 5th IFAC Symposium on Mechatronic Systems, Boston

Proakis JG, Manolakis DG (1996) Digital Signal Processing. Prentice Hall

Qin SJ, Badgwell TA (1996) An overview of industrial model predictive control technology, preprint CPC–V, Lake Tahoe

Qin SJ, Badgwell TA (2000) An overview of nonlinear model predictive control applications. In: Nonlinear Predictive Control, Springer–Verlag, pp 369—-392

Qu CC, Hahn J (2009) Computation of arrival cost for moving horizon estimation via unscented Kalman filtering. J Process Control 19:358–363

Raff T, Ebenbauer C, Findeisen R, Allgöwer F (2005) Remarks on moving horizon state estimation with guaranteed convergence. In: Meurer T, Graichen K, Gilles ED (eds) Control and Obsever Design for Nonlinear Finite and Infinite Dimensional Systems, Springer–Verlag, Berlin, pp 67–80

Ramamurthi Y, Sistu P, Bequette BW (1993) Control-relevant data reconciliation and parameter estimation. Comp Chem Engr 17:41–59

Rao CV, Rawlings JB, Mayne DQ (2003) Constrained state estimation for nonlinear discrete–time systems: Stability and moving horizon approximation. IEEE Transactions Automatic Control 48:246–258

Rawlings JB (2000) Tutorial overview of model predictive control. IEEE Contr Syst Magazine 20(3):38–52

Rawlings JB, Bakshi BR (2006) Particle filtering and moving horizon estimation. Computers and Chemical Engineering 30:1529–1541

Roset BJP, Lazar M, Heemels WPMH, Nijmeijer H (2006) A stabilizing output based nonlinear model predictive control scheme. In: Proceedings of 45th IEEE Conference on Decision and Control, San Diego, USA, pp 4627–4632

Saberi A, Han J, Stoorvogel AA (2002) Constrained stabilization problems for linear plants. Automatica 38:639–654

Sargent RWH, Sullivan GR (1977) The development of an efficient optimal control package. In: Proceedings of the 8th IFIP Conference on Optimization Techniques, Springer, Heidelberg

Scattolini R (2009) Architectures for distributed and hierarchical model predictive control – a review. Journal of Process Control 19:723–731
Schei TS (1997) A finite-difference method for linearization in nonlinear estimation algorithms. Automatica 33:2053–2058
Scokaert POM, Mayne DQ, Rawlings JB (1999) Suboptimal model predictive control (feasibility implies stability). IEEE Trans Automatic Control 44:648–654
Sepulchre R, Jankovic M, Kokotovic P (1997) Constructive nonlinear control. Springer–Verlag, London
Sequeira SE, Graells M, Luis P (2002) Real-time evolution for on-line optimization of continuous processes. Industrial and Engineering Chemistry Research pp 1815–1825
Shahzad A, Kerrigan EC, Constantinides GA (2010) A warm-start interior-point method for predictive control. In: Proc. UKACC
Shin KG, McKay ND (1985) Minimum-time control of robotic manipulators with geometric path constraints. IEEE Transactions on Automatic Control 30:531–541
von Stryk O (1993) Numerical solution of optimal control problems by direct collocation. In: in Optimal Control, (International Series in Numerical Mathematics 111), pp 129–143
Sui D, Johansen TA (2010) Regularized nonlinear moving horizon observer for detectable systems. In: IFAC Nonlinear Control Symposium, Bologna, Italy
Sui D, Johansen TA, Feng L (2010) Linear moving horizon estimation with pre-estimating observer. IEEE Trans Automatic Control 55:2363 – 2368
Tenny MJ, Rawlings JB, Wright SJ (2004) Closed-loop behavior of nonlinear model-predictive control. AIChE Journal 50:2142–2154
Tikhonov AN, Arsenin VY (1977) Solutions of Ill–posed Problems. Wiley
Tjoa IB, Biegler LT (1991) Simultaneous strategies for data reconciliation and gross error detection for nonlinear systems. Comp Chem Engr 15:679–690
Tsang TH, Himmelblau DM, Edgar TF (1975) Optimal control via collocation and non–linear programming. Int J Control 21:763–768
Tyler ML, Morari M (1999) Propositional logic in control and monitoring problems. Automatica 35:565–582
Ungarala S (2009) Computing arrival cost parameters in moving horizon estimation using sampling based filters. J Process Control 19:1576–1588
Vada J, Slupphaug O, Foss BA (1999) Infeasibility handling in linear MPC subject to prioritized constraints. In: IFAC World Congress, Beijing, vol D, pp 163–168
Wan ZY, Kothare MV (2004) Efficient scheduled stabilizing output feedback model predictive control for constrained nonlinear systems. IEEE Trans Automatic Control 49:1172 – 1177
Williams HP (1999) Model Building in Mathematical Programming
Wright SJ (1997) Primal-dual interior-point methods. SIAM, Philadelphia
Wright SJ, Tenny MJ (2004) A feasible trust-region sequential quadratic programming algorithm. SIAM Journal on Optimization 14:1074–1105
Xiong Q, Jutan A (2003) Continuous optimization using a dynamic simplex method. Chemical Engineering Science pp 3817–2828
Zavala VM, Biegler LT (2009) The advanced step NMPC controller: Optimality, stability and robustness. Automatica 45:86–93
Zavala VM, Laird CD, Biegler LT (2008) A fast moving horizon estimation algorithm based on nonlinear programming sensitivity. Journal of Process Control 18:876–884
Zimmer G (1994) State observation by on–line minimization. Int J Control 60:595–606