Constrained State Estimation for Nonlinear Discrete-Time Systems: Stability and Moving Horizon Approximations

Christopher V. Rao, James B. Rawlings, and David Q. Mayne

***Abstract*—State estimator design for a nonlinear discrete-time system is a challenging problem, further complicated when additional physical insight is available in the form of inequality constraints on the state variables and disturbances. One strategy for constrained state estimation is to employ online optimization using a moving horizon approximation. In this article we propose a general theory for constrained moving horizon estimation. Sufficient conditions for asymptotic and bounded stability are established. We apply these results to develop a practical algorithm for constrained linear and nonlinear state estimation. Examples are used to illustrate the benefits of constrained state estimation. Our framework is deterministic.**

***Index Terms*—Constraints, model predictive control (MPC), moving horizon estimation (MHE), optimization, state estimation.**

I. Introduction

OUR problem concerns the design of constrained state estimators for nonlinear discrete-time systems, where one possesses additional insights in the form of general inequality constraints on the state variables and disturbances. Constraints are typically used to model bounded disturbances, though they are also used to correct for model error by bounding the state. While many powerful strategies exist for estimating the state of nonlinear discrete-time systems, these strategies do not address the issue of constraints.

The constrained state estimation problem can be reformulated as a series of optimal control problems (cf. [1] and [2]). Solving the optimal control problems, however, is computationally demanding, because the problem dimension grows with time as more data are processed. One method to reduce the computational burden is to bound the size of the estimation problem by employing a moving horizon approximation. Moving horizon approximations have been used successfully to develop stabilizing control laws for constrained nonlinear systems (cf. [3]). In moving horizon estimation (MHE), the state estimate is determined online by solving a finite horizon state estimation problem. As new measurements become available, the old measurements are discarded from the estimation window, and the finite horizon state estimation problem is resolved to determine the new estimate of the state. The method is optimization based, so MHE can handle explicitly nonlinear systems and inequality constraints on the decision variables.

In this paper, we investigate online optimization strategies for estimating the state of systems modeled by a nonlinear difference equation of the form

$$\begin{aligned} x_{k+1} &= f_k(x_k, w_k) \\ y_k &= h_k(x_k) + v_k \end{aligned} \tag{1}$$

where it is known that the states and disturbances satisfy the following constraints:

$$x_k \in \mathbb{X}_k \quad w_k \in \mathbb{W}_k \quad v_k \in \mathbb{V}_k.$$

We assume, for all $k \geq 0$, the functions $f_k : \mathbb{X}_k \times \mathbb{W}_k \to \mathbb{X}_k$ and $h_k : \mathbb{X}_k \to \mathbb{R}^p$ and the sets $\mathbb{X}_k \subseteq \mathbb{R}^n$, $\mathbb{W}_k \subseteq \mathbb{R}^m$, and $\mathbb{V}_k \subseteq \mathbb{R}^p$ are closed with $0 \in \mathbb{W}_k$ and $0 \in \mathbb{V}_k$.

Let $x(k; z, l, \{w_j\})$ denote the solution of the system (1) at time k when the initial state is z at time l and the input disturbance sequences is $\{w_j\}_{j=l}^{k}$. When we consider the disturbance free response of the system, i.e., $\{w_j\} = \{0\}$, we use the following notational simplification $x(k; z, l)$. Let $y(k; z, l, \{w_j\}) := h_k(x(k; z, l, \{w_j\}))$ denote the output response of the system (1) at time k when the initial state is z at time l and the input disturbance sequences is $\{w_j\}_{j=l}^{k}$. We use the notational simplification $y(k; z, l) := h_k(x(k; z, l))$ for the disturbance free output response of the system. Note the difference between y_k and $y(k; z, l, \{w_j\})$. The vector y_k denotes the observed output at time k and the vector $y(k; z, l, \{w_j\})$ denotes the predicted output at time k when the initial condition at time l is z and the disturbance sequence is $\{w_j\}_{j=l}^{k}$.

One may interpret the constraints $\mathbb{W}_k$ and $\mathbb{V}_k$ as a strategy for modeling bounded disturbances or random variables with truncated densities. However, the interpretation of the state constraints $\mathbb{X}_k$ is not so simple. If the state is subject to physical constraints such as concentrations are positive, then the constraints should be satisfied implicitly by the model (1). However, if the physical constraints are not implicitly enforced by the model, then state constraints may be used to account for model inaccuracies. In particular, state constraints may be used to simplify the model. Thus, state constraints are nonstandard; one usually

Manuscript received September 7, 2001; revised May 29, 2002 and August 28, 2002. Recommended by Associate Editor T. Parisini. This work was supported by the Industrial Members of the Texas-Wisconsin Modeling and Control Consortium, and by the National Science Foundation under Grant CTS-9708497.

C. V. Rao is with the Department of Bioengineering, University of California, Berkeley, CA 94720 USA (e-mail: c_rao@lbl.gov).

J. B. Rawlings is with the Department of Chemical Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: jbraw@che.wisc.edu).

D. Q. Mayne is with the Department of Electrical & Electronic Engineering, Imperial College of Science, Technology, and Medicine,SW7 2BT London, U.K. (e-mail: d.mayne@ic.ac.uk).

Digital Object Identifier 10.1109/TAC.2002.808470

chooses an exact model of the plant and, separately, the characteristics of the disturbances, such as boundedness, or that the disturbances are independent and identically distributed with known mean and variance. The properties of the model and disturbances are distinct. State constraints, on the other hand, correlate the disturbances with the state and may lead to acausality. While not always theoretically satisfying, state constraints may be appealing to the engineer. The issues regarding state constraints have not been resolved completely. For further discussion, the reader is directed to [4].

We formulate the constrained estimation problem, for $T \geq 0$, as the solution to the following optimal control problem:

$$P_1(T): \quad \Phi_T^* = \min_{x_0, \{w_k\}_{k=0}^{T-1}} \{\Phi_T(x_0, \{w_k\}) : (x_0, \{w_k\}) \in \Omega_T\}$$

where the objective function is defined by

$$\Phi_T(x_0, \{w_k\}) := \sum_{k=0}^{T-1} L_k(w_k, v_k) + \Gamma(x_0)$$

the constraint set is defined by the equation shown at the bottom of the page, and $v_k := y_k - y(k; x_0, 0, \{w_j\})$. We assume the stage cost function $L_k : \mathbb{W}_k \times \mathbb{V}_k \to \mathbb{R}_{\geq 0}$ for all $k \geq 0$ and the initial penalty $\Gamma : \mathbb{X}_0 \to \mathbb{R}_{\geq 0}$. The initial penalty $\Gamma(\cdot)$ summarizes the prior information at time $k = 0$ and satisfies $\Gamma(\hat{x}_0) = 0$, where $\hat{x}_0 \in X_0$ is the *a priori* most likely value of x_0, and $\Gamma(x) > 0$ for $x \neq \hat{x}_0$; The initial penalty $\Gamma(\cdot)$ is part of the data of the state estimation problem. Typically

$$\Gamma(x) := (x - \hat{x}_0)^T \overline{\Pi}_0^{-1} (x - \hat{x}_0)$$

where the matrix $\overline{\Pi}_0$ is symmetric positive definite. In this case, the given data $(\hat{x}_0, \overline{\Pi}_0)$ determines $\Gamma(\cdot)$. The solution to $P_1(T)$ at time T is the pair

$$\left(\hat{x}_{0|T-1}, \{\hat{w}_{k|T-1}\}_{k=0}^{T-1}\right)$$

and that optimal pair yields an estimate $\{\hat{x}_{k|T-1}\}_{k=0}^{T}$ of the actual sequence $\{x_k\}$; the sequence $\{\hat{x}_{k|T-1}\}_{k=0}^{T}$ is the solution of (1) with the initial state $\hat{x}_{0|T-1}$ at time $k = 0$ and disturbance sequence $\{\hat{w}_{k|T-1}\}_{k=0}^{T-1}$, i.e.,

$$\hat{x}_{k|T-1} := x(k; \hat{x}_{0|T-1}, 0, \{\hat{w}_{j|T-1}\}).$$

To simplify notation $\hat{x}_j := \hat{x}_{j|j-1}$, where $\hat{x}_{0|-1} = \hat{x}_0$.

We refer to the formulation $P_1(T)$ as the full information problem and $\hat{x}_k$ as the full information estimate of x_k, because all of the available information $\{y_k\}_{k=0}^{T-1}$ is processed. The problem $P_1(T)$ has T stages, so the computational complexity scales at least linearly with T. Unless the process is linear, unconstrained, and the cost functions are quadratic, in which case the optimal estimator is the Kalman filter and the solution is obtained recursively, the online solution of $P_1(T)$ is impractical because the computational burden increases with time. To make the problem tractable, we need to bound the problem size. One strategy to reduce $P_1(T)$ to a fixed-dimension optimal control problem is to employ a moving horizon approximation. Unlike the full-information problem, MHE does not estimate the full-state sequence $\{x_k\}_{k=0}^{T}$. Rather, MHE estimates the truncated sequence $\{x_k\}_{k=T-N}^{T}$. The key to preserving stability and performance is how one approximately summarizes the past data.

Consider again the problem $P_1(T)$. We can arrange the objective function by breaking the time interval into two pieces as follows:

$$\Phi_T(x_0, \{w_k\}) = \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \sum_{k=0}^{T-N-1} L_k(w_k, v_k) + \Gamma(x_0).$$

Because we use a state-variable description of the system, the quantity

$$\sum_{k=T-N}^{T-1} L_k(w_k, v_k)$$

depends only on the state x_{T-N} and the sequences $\{w_k, v_k\}_{k=T-N}^{T-1}$. Exploiting the relation using forward dynamic programming, we can establish the equivalence between a full information problem and an estimation problem with a fixed-size estimation window.

Consider the reachable set of states at time τ generated by a feasible initial condition x_0 and disturbance sequence $\{w_k\}_{k=0}^{\tau-1}$

$$\mathcal{R}_\tau = \{x(\tau; x_0, 0, \{w_j\}) : (x_0, \{w_j\}) \in \Omega_\tau\}.$$

where Ω_τ is defined below. We define the **arrival cost**[1] at time τ and for the state $z \in \mathcal{R}_\tau$ as

$$\mathcal{Z}_\tau(z) := \min_{x_0, \{w_k\}_{k=0}^{\tau-1}} \{\Phi_\tau(x_0, \{w_k\}) : (x_0, \{w_k\}) \in \Omega_\tau,\ x(\tau; x_0, 0, \{w_j\}) = z\}.$$

It follows that $\mathcal{Z}_0(\cdot) = \Gamma(\cdot)$. Arrival cost is a fundamental concept in MHE, because we can reformulate $P_1(T)$, for $T > N$, as the following equivalent optimal control problem:

$$P_1'(T) \quad \Phi_T^* = \min_{z, \{w_k\}_{k=T-N}^{T-1}} \left\{ \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \mathcal{Z}_{T-N}(z) : \begin{array}{c} z \in \mathcal{R}_{T-N} \\ (z, \{w_k\}) \in \Omega_T^N \end{array} \right\}$$

[1]Other researchers have used the term **cost to come** (cf. [5]) or **cost to arrive** (cf. [6]).

$$\Omega_T := \left\{ (x_0, \{w_k\}) : \begin{array}{ll} x(k; x_0, 0, \{w_j\}) \in \mathbb{X}_k, & k = 0, \ldots T \\ w_k \in \mathbb{W}_k, & k = 0, \ldots, (T-1) \\ v_k = y_k - y(k; x_0, 0, \{w_j\}) \in \mathbb{V}_k, & k = 0, \ldots, (T-1) \end{array} \right\}$$

where the constraint set is defined the equation shown at the bottom of the page, and $v_k := y_k - y(k; z, T-N, \{w_k\})$. When $T \leq N$, the optimal control problem $P_1'(T)$ is defined to be $P_1(T)$. It is relatively straightforward to demonstrate the equivalence of the solutions to $P_1(T)$ and $P_1'(T)$ using forward dynamic programming (cf. [7]).

Optimality guarantees $\mathcal{Z}_T(z) \geq \Phi_T^*$ for all $z \in \mathcal{R}_T$ and $\mathcal{Z}_T(\hat{x}_T) = \Phi_T^*$. We can view, therefore, the arrival cost as an equivalent statistic [8] for summarizing the past data $\{y_k\}_{k=0}^{T-N-1}$ not explicitly accounted for in the objective function of $P_1'(T)$. The arrival cost serves as an equivalent statistic by penalizing the deviation of x_{T-N} away from $\hat{x}_{T-N}$. If we have high (low) confidence in the optimal estimate $\hat{x}_{T-N}$, then the cost of choosing x_{T-N} far away from $\hat{x}_{T-N}$ is large (small).

For the majority of systems, we do not possess algebraic expressions for the arrival cost. Notable exceptions are unconstrained linear systems with quadratic objectives, where the estimate $\hat{x}_j$ is now the standard Kalman estimate of the state x_j. Assume the functions $f_k(\cdot)$ and $h_k(\cdot)$ are defined by

$$f_k(x,w) := A_k x + G_k w \quad h_k(x) := C_k x$$

and the stage penalties $L_k(\cdot)$ are defined by

$$L_k(w,v) := w^T Q_k^{-1} w + v^T R_k^{-1} v$$

where the matrices Q_k and R_k are symmetric positive definite. For this case, the initial penalty is defined as

$$\Gamma(x) := (x - \hat{x}_0)^T \overline{\Pi}_0^{-1} (x - \hat{x}_0)$$

and the arrival cost is given by

$$\mathcal{Z}_j(z) := (z - \hat{x}_j)^T \Pi_j^{-1} (z - \hat{x}_j) + \Phi_j^* \tag{2}$$

assuming the matrix Π_j is invertible. The matrix sequence $\{\Pi_j\}$ is obtained by solving the matrix Riccati equation

$$\begin{aligned}\Pi_{j+1} = G_j Q_j G_j^T + A_j \Pi_j A_j^T \\ - A_j \Pi_j C_j^T (R_j + C_j \Pi_j C_j^T)^{-1} C_j \Pi_j A_j^T\end{aligned} \tag{3}$$

with the initial condition $\Pi_0 = \overline{\Pi}_0$. One obtains this result by deriving the deterministic Kalman filter using forward dynamic programming (cf. [1]).

When the system is nonlinear or constrained, an algebraic expression for the arrival cost rarely exists, yet we require one to successfully implement the estimator. Ideally, we want the moving horizon estimate as close as possible to the full information estimate. One solution is to formulate MHE as the solution to a numerically tractable though approximate version of $P_1'(T)$. An approximation $\hat{\mathcal{Z}}_j(\cdot)$ of the arrival cost $\mathcal{Z}_j(\cdot)$ may be used to account for the data not included in the estimation window. The past data are accounted for approximately with our choice of $\hat{\mathcal{Z}}_j(\cdot)$ by penalizing deviation away from the past estimate $\hat{x}_j$ in accordance with our confidence in the estimate. Because this choice is an approximation, we need to ensure that estimator divergence does not result. In examples not shown here (see [4]), we demonstrate how poor approximations of the arrival cost may lead to estimator divergence. In Section III, we discuss the stability implications of approximate representations of the arrival cost.

We formulate, for $T > N$, the moving horizon approximation to the full information estimation problem, or MHE, as the following optimal control problem:

$$P_2(T) \quad \hat{\Phi}_T = \min_{z, \{w_k\}_{k=T-N}^{T-1}} \left\{ \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \hat{\mathcal{Z}}_{T-N}(z) : (z, \{w_k\}) \in \Omega_T^N \right\}$$

where $v_k := y_k - y(k; z, T-N, \{w_j\})$ and $\hat{\mathcal{Z}}_j : \mathbb{X}_j \to \mathbb{R}$ for all $j \geq 0$. The moving horizon cost $\hat{\Phi}_T$ approximates Φ_T^* by replacing the (uncomputable) arrival cost $\mathcal{Z}_{T-N}(\cdot)$ with an approximation $\hat{\mathcal{Z}}_{T-N}(\cdot)$ and removing the constraint $z \in \mathcal{R}_{T-N}$. We choose $\hat{\mathcal{Z}}_0(\cdot) = \Gamma(\cdot)$. When $T \leq N$, the optimal control problem $P_2(T)$ is defined to be $P_1(T)$. The solution to $P_2(T)$ at time T is the pair $(z^*, \{\hat{w}_{k|T-1}^{\text{mh}}\}_{k=T-N}^{T-1})$, which, when used as data in the system (1), yields $\{\hat{x}_{k|T-1}^{\text{mh}}\}_{k=T-N}^{T}$, i.e.,

$$\hat{x}_{k|T-1}^{\text{mh}} := x\left(k; z^*, T-N, \left\{\hat{w}_{j|T-1}^{\text{mh}}\right\}\right).$$

For simplicity, $\hat{x}_j^{\text{mh}} := \hat{x}_{j|j-1}^{\text{mh}}$, where $\hat{x}_{0|-1}^{\text{mh}} = \hat{x}_0$.

One strategy to approximate the arrival cost $\mathcal{Z}_T(\cdot)$ is to employ a first-order Taylor series approximation of the model (1) around the estimated trajectory $\{\hat{x}_k^{\text{mh}}\}_{k=0}^{T}$. This strategy yields an extended Kalman filter covariance update formula for constructing $\hat{\mathcal{Z}}_T(\cdot)$. Suppose the model functions $f_k(\cdot)$ and $h_k(\cdot)$ and the cost functions $L_k(\cdot)$ are sufficiently smooth and

$$\Gamma(x) := (x - \hat{x}_0)^T \overline{\Pi}_0^{-1} (x - \hat{x}_0).$$

Let

$$A_k := \left.\frac{\partial f_k(x,0)}{\partial x}\right|_{\hat{x}_k^{\text{mh}}} \quad G_k := \left.\frac{\partial f_k(\hat{x}_k^{\text{mh}}, w)}{\partial w}\right|_{w=0}$$
$$C_k := \left.\frac{\partial h_k(x)}{\partial x}\right|_{\hat{x}_k^{\text{mh}}}$$

denote the linearized dynamics of (1) and

$$R_k^{-1} := \left.\frac{\partial^2 \underline{L}_k(0,v)}{\partial v \partial v^T}\right|_{\hat{x}_k^{\text{mh}}} \quad N_k := \left.\frac{\partial^2 \underline{L}_k(w,v)}{\partial w \partial v^T}\right|_{w=0,\ \hat{x}_k^{\text{mh}}}$$
$$Q_k^{-1} := \left.\frac{\partial^2 \underline{L}_k(w,v)}{\partial w \partial w^T}\right|_{w=0,\ \hat{x}_k^{\text{mh}}}$$

$$\Omega_T^N := \left\{ (z, \{w_k\}) : \begin{array}{cc} x(k; z, T-N, \{w_j\}) \in \mathbb{X}_k, & k = (T-N), \ldots, T \\ w_k \in \mathbb{W}_k, & k = (T-N), \ldots, (T-1) \\ v_k = y_k - y(k; z, T-N, \{w_j\}) \in \mathbb{V}_k, & k = (T-N), \ldots, (T-1) \end{array} \right\}$$

denote the linearized stage penalties $L_k(\cdot)$, then, if we assume for simplicity $N_k = 0$, we approximate the arrival cost as

$$\hat{\mathcal{Z}}_T(z) = \left(z - \hat{x}_T^{\text{mh}}\right)^T \Pi_T^{-1} \left(z - \hat{x}_T^{\text{mh}}\right) + \hat{\Phi}_T$$

assuming the matrix Π_T is invertible, where the matrix sequence $\{\Pi_j\}$ is obtained by solving the matrix Riccati (3) subject to the initial condition $\Pi_0 = \overline{\Pi}_0$. This result is equivalent to the covariance update formula for the extended Kalman filter. See [2] for further details.

MHE is a practical strategy to handle the computational difficulties associated with optimization based estimation, and, as a consequence, many authors have explored different issues in MHE. The first application of MHE for nonlinear systems was the work of Jang *et al.* [9]. Their strategy ignores disturbances and constraints and attempts to estimate only the initial state of the system. Thomas [10] and Kwon *et al.* [11] discussed earlier moving horizon strategies for unconstrained linear systems. Limited memory and adaptive filters for linear systems are analogous to MHE, because only a fixed window of data is considered (see [2] for a discussion of limited memory filters). Many researchers in the process systems area extended the work of Jang *et al.*. Bequette *et al.* [12], [13] investigated moving horizon strategies for state estimation as a logical extension of model predictive control. Edgar and coworkers [14], [15] investigated moving horizon strategies for nonlinear data reconciliation. Biegler *et al.* [16]–[18] investigated statistical and numerical issues related to optimization-based nonlinear data reconciliation. Marquardt *et al.* [19], [20] discussed multi-scale strategies for MHE and the benefits of incorporating constraints in estimation. Bemporad *et al.* [21] discussed the application of MHE to hybrid systems. Gesthuisen and Engell [22] discussed the application of MHE to a pilot-scale polymerization reactor and Russo and Young [23] discussed the application of MHE to an industrial polymerization process at the Exxon Chemical Company. Because MHE is formulated as an optimization problem, it is possible to handle explicitly inequality constraints. Robertson and Lee [24]–[26] have investigated the probabilistic interpretation of constraints in estimation. Muske and Rawlings [27], [28] derived some preliminary conditions for the stability of state estimation with inequality constraints. Tyler and Morari [29], [30] demonstrated how constraints may result in instability for nonminimum phase systems.

In parallel to the research done in process systems, unconstrained MHE was investigated also by researchers in automatic control. Ling and Lim [31] and Kwon *et al.* [32], [33] investigated the MHE for linear systems. Zimmer [34] investigated an unconstrained MHE strategy for nonlinear systems similar to the approach of Jang and coworkers [9] and derived conditions for stability using fixed point theorems. Moraal and Grizzle [35] also derived conditions for stability for nonlinear MHE using fixed point theorems. However, Moraal and Grizzle [35] formulated the estimation problem as the solution to a set of algebraic equations. Michalska and Mayne [36] investigated an unconstrained MHE strategy for nonlinear systems similar to the approach of Jang *et al.* [9] and derived conditions for stability using Lyapunov arguments. Vincent and Khargonekar [37] investigated unconstrained MHE for a class of systems arising from drifting sensor gains. Alessandri *et al.* [38] investigated MHE for systems with bounded measurement error and developed error bounds on the resulting estimates. Our results are novel in that we investigate the stability properties of MHE under general constraints on the state and disturbances.

The remainder of the paper is organized as follows. Section II introduces the notation, definitions, and basic assumptions necessary for establishing stability. We establish sufficient conditions for the asymptotic and bounded stability of MHE and propose a prototype algorithm for MHE in Section III. We conclude in Section IV by illustrating the effectiveness of MHE for constrained estimation with numerical examples. Extensions of the results presented, including discussions of duality and suboptimality, are available in [4].

II. Notation, Definitions, and Basic Assumptions

The Cartesian product $\times_{k=1}^{N}\mathbb{A}$ of a set $\mathbb{A}$ is denoted by $\mathbb{A}^N$. We use the symbol $\|\cdot\|$ to denote any vector norm in $\mathbb{R}^n$ (where the dimension n follows from context). Let $\mathbb{R}_{\geq 0}$ denote the nonnegative real numbers, and $\underline{\mathcal{C}}(\mathbb{R}^n)$ denote the space of lower semi-continuous functions that map from $\mathbb{R}^n$ to $\mathbb{R}$. For $\epsilon > 0$, $B_\epsilon := \{x : \|x\| \leq \epsilon\}$. For notational simplicity, we make that following definition: $\hat{(\cdot)}_k := \hat{(\cdot)}_{k|k-1}$.

Definition 2.1: A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a **K-function** if it is continuous, strictly monotone increasing, $\alpha(x) > 0$ for $x \neq 0$, $\alpha(0) = 0$, and $\lim_{x\to\infty} \alpha(x) = \infty$.

Throughout this paper, we use the following elementary properties of K-functions.

Fact 2.2: Suppose $\alpha(\cdot)$ is a K-function. Then, the function $\alpha^{-1} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a K-function [39].

Fact 2.3: The space of K-functions is closed under addition, composition, and positive scalar multiplication. For example, if $\alpha(\cdot)$ and $\beta(\cdot)$ are K-functions, then $\alpha \circ \beta(\cdot)$, $\alpha(\cdot) + \beta(\cdot)$, $c\alpha(\cdot)$ for $c > 0$ are K-functions.

To establish existence and stability, we require the following observability condition.

Definition 2.4: A system is **uniformly observable** if there exists a positive integer N_o and a K-function $\varphi(\cdot)$ such that for any two states x_1 and x_2

$$\varphi(\|x_1 - x_2\|) \leq \sum_{j=0}^{N_o-1} \|y(k+j; x_1, k) - y(k+j; x_2, k)\|$$

for all $k \geq 0$.

In order to guarantee the problems $P_1(T)$ and $P_2(T)$ are well posed, we require that the model (1), stage cost functions $L_k(\cdot)$, and initial penalty $\Gamma(\cdot)$ satisfy the following conditions.

A0) The functions $f_k(\cdot)$ and $h_k(\cdot)$ are Lipschitz continuous in all of their arguments with constants c_f and c_h respectively for all $k \geq 0$.

A1) $L_k(\cdot) \in \underline{\mathcal{C}}(\mathbb{W}_k \times \mathbb{V}_k)$ for all $k \geq 0$ and $\Gamma(\cdot) \in \underline{\mathcal{C}}(\mathbb{X}_0)$.

A2) There exist K-functions $\eta(\cdot)$ and $\gamma(\cdot)$ such that

$$\eta(\|(w,v)\|) \leq L_k(w,v) \leq \gamma(\|(w,v)\|)$$
$$\eta(\|x - \hat{x}_0\|) \leq \Gamma(x) \leq \gamma(\|x - \hat{x}_0\|)$$

for all $(w,v) \in (\mathbb{W}_k \times \mathbb{V}_k)$, $x \in \mathbb{X}_0$, $\hat{x}_0 \in \mathbb{X}_0$, and $k \geq 0$.

Assumption **A0)** is satisfied if the functions $f_k(\cdot)$ and $h_k(\cdot)$ are twice differentiable. Assumptions **A1)** and **A2)** are satisfied if the stage cost functions $L_k(\cdot)$ and the initial penalty $\Gamma(\cdot)$ are positive definite quadratic functions.

We need also to impose similar conditions on the approximate arrival cost $\hat{\mathcal{Z}}_k(\cdot)$. However, unlike the initial penalty, the minimal value of the arrival cost is greater than zero (recall $\mathcal{Z}_k(x) \geq \hat{\Phi}_k$ for all $x \in \mathbb{X}_k$ with $\mathcal{Z}_k(\hat{x}_k) = \hat{\Phi}_k$) and the approximate arrival $\hat{\mathcal{Z}}_k(x)$ may not be bounded below by $\|x\|$ for reasons that become apparent in Section III. We require instead $\hat{\mathcal{Z}}_k(\cdot)$ satisfies the following condition.

C1) There exist K-function $\bar{\gamma}(\cdot)$ such that

$$0 \leq \hat{\mathcal{Z}}_k(z) - \hat{\Phi}_k \leq \bar{\gamma}(\|z - \hat{x}_k^{\mathrm{mh}}\|)$$

for all $z \in \mathbb{X}_T$, and $T \geq 0$.

The following technical lemma follows from the definition of observability and definitions previously given.

Lemma 2.5: Suppose **A0)–A2)** are true and (1) is uniformly observable. If the exists positive constants δ_w and δ_v such that $\mathbb{W}_k \subseteq B_{\delta_w}$ and $\mathbb{V}_k \subseteq B_{\delta_v}$ for all k, then, for all $N \geq N_o$, there exists a K-function $\theta(\cdot)$ such that

$$\|x_i - \hat{x}_i\| \leq \theta\left(\left\|\left(\left\{\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1}, \hat{v}_{k|T-1})\right\}, \delta_w, \delta_v\right)\right\|\right)$$

for $T \geq N$ and all $i \in \{T-N, T-N+1, \ldots, T\}$.

Proof: The proof is given in Appendix A. □

To guarantee that a solution exists to either $P_1(T)$ or $P_2(T)$, we require that the feasible region is nonempty.

A3) There exists an initial condition $x_{0|\infty}$, disturbance sequence $\{w_{k|\infty}\}_{k=0}^{\infty}$ such that, for all $k \geq 0$, $\left(x_{0|\infty}, \{w_{k|\infty}\}\right) \in \Omega_k$.

To account for constraints, we have modified slightly the definition of stability in an analogous manner to [40].

Definition 2.6: An estimator is an **asymptotically stable observer** for the system

$$x_{k+1} = f_k(x_k, 0) \quad y_k = h_k(x_k) \tag{4}$$

if, for every initial condition x_0 such that $x(k; x_0, 0) \in \mathbb{X}_k$ for all $k \geq 0$ and every $\epsilon > 0$, there corresponds a number $\delta > 0$ and a positive integer $\bar{T}$ such that if $\|x_0 - \hat{x}_0\| \leq \delta$ and $\hat{x}_0 \in \mathbb{X}_0$, then $\|x(T; x_0, 0) - \hat{x}_T\| \leq \epsilon$ for all $T \geq \bar{T}$. Furthermore, for all $x_0 \in \mathbb{X}_0$, $\hat{x}_T \to x(T; x_0, 0)$ as $T \to \infty$.

III. Stability

In this section we derive sufficient conditions for asymptotic and bounded stability. We begin by stating conditions on the approximate arrival cost $\hat{\mathcal{Z}}_j(\cdot)$ sufficient to guarantee the stability of MHE. We proceed to derive conditions for the existence of a solution to $P_2(T)$, and we then establish stability. For most nonlinear systems the approximate arrival costs are unable to satisfy *a priori* the stability condition. We then present an algorithm for constrained MHE that relaxes the conditions on the approximate arrival costs. We conclude the section by establishing bounded stability in the presence of bounded noise.

Ideally the approximate arrival cost $\hat{\mathcal{Z}}_j(\cdot)$ is equal to the arrival cost $\mathcal{Z}_j(\cdot)$. With the notable exception of the unconstrained linear quadratic problem (i.e., the Kalman filter), closed-form expressions for the arrival cost are generally unavailable. To guarantee stability, however, we do not need to construct the arrival cost, but rather require instead that the approximate arrival cost satisfies the following condition.

C2) Let

$$\mathcal{R}_\tau^N = \left\{\, x(\tau; z, \tau - N, \{w_k\}) : (z, \{w_k\}) \in \Omega_\tau^N \right\}$$

where $\mathcal{R}_\tau^N = \mathcal{R}_\tau$ for $\tau \leq N$. For a horizon length N, any time $\tau > N$, and any $p \in \mathcal{R}_\tau^N$, the approximate arrival cost $\hat{\mathcal{Z}}_\tau(\cdot)$ satisfies the inequality

$$\hat{\mathcal{Z}}_\tau(p) \leq \min_{z, \{w_k\}_{k=\tau-N}^{\tau-1}} \left\{ \sum_{k=\tau-N}^{\tau-1} L_k(w_k, v_k) + \hat{\mathcal{Z}}_{\tau-N}(z) : \begin{matrix} (z, \{w_k\}) \in \Omega_\tau^N, \\ x(\tau; z, \tau-N, \{w_j\}) = p \end{matrix} \right\} \tag{5}$$

subject to the initial condition $\hat{\mathcal{Z}}_0(\cdot) = \Gamma(\cdot)$. For $\tau \leq N$, the approximate arrival cost $\hat{\mathcal{Z}}_\tau(\cdot)$ satisfies instead the inequality $\hat{\mathcal{Z}}_\tau(\cdot) \leq \mathcal{Z}_\tau(\cdot)$.

If one views arrival cost as an equivalent statistic for the data, then the inequality (5) in condition **C2)** states that the approximate arrival cost should not add additional "information" not specified in the data. Loosely speaking, we say a positive function $a(\cdot)$ contains more information than another positive function $b(\cdot)$ if $a(x) \geq b(x)$ for all x of interest. If the inequality (5) were strict, then condition **C2)** would state there should be some "forgetting" in the estimator.

Remark 3.1: A simple strategy to satisfy condition **C2)** is to define for time τ the approximate arrival cost as $\hat{\mathcal{Z}}_\tau(\cdot) := \hat{\Phi}_\tau$. The inequality (5) is satisfied by definition: optimality of $P_2(\tau)$ guarantees that the optimal cost $\hat{\Phi}_\tau$ satisfies the inequality (5) for all $p \in \mathcal{R}_\tau^N$. This construction was employed by Muske and Rawlings [28] to generate a stable nonlinear MHE. Without constraints, this choice yields a deadbeat observer.

Remark 3.2: If we choose

$$\hat{\mathcal{Z}}_\tau(z) = \left(z - \hat{x}_\tau^{mh}\right)^T \Pi_\tau^{-1} \left(z - \hat{x}_\tau^{mh}\right) + \hat{\Phi}_\tau$$

where the sequence $\{\Pi_j\}$ is obtained by solving the matrix Riccati (3) subject to the initial condition $\Pi_0 = \overline{\Pi}_0$,, then condition **C2)** is satisfied when we consider linear systems with quadratic objectives and convex constraints. The proof of this claim is given in [4].

We begin by providing sufficient conditions for the existence of a solution to $P_2(T)$.

Proposition 3.3: If assumptions **A0)–A3)** hold, the sequence $\{\hat{\mathcal{Z}}_j(\cdot)\}$ satisfies condition **C1)**, the system (4) is uniformly observable, and $N \geq N_o$, then a solution exists to $P_2(T)$ for all $\hat{x}_0 \in \mathbb{X}_0$ and $T \geq 0$.

Proof: The proof is given in Appendix B. □

In the following proposition, we state our stability result for MHE. Stability is established by demonstrating that the sequence $\{\hat{\Phi}_k\}$ is nondecreasing and bounded above uniformly for $k \geq 0$ by the initial estimation error $\|x_0 - \hat{x}_0\|$.

Proposition 3.4: If assumptions **A0)–A3)** hold, the sequence $\{\hat{\mathcal{Z}}_j(\cdot)\}$ satisfies conditions **C1)** and **C2)**, system (1) is uniformly observable, and $N \geq N_o$, then, for all $\hat{x}_0 \in \mathbb{X}_0$, MHE is an asymptotically stable observer for the system (4).

Proof: We first prove convergence by demonstrating that $\gamma(\|x_0 - \hat{x}_0\|)$, where $\gamma(\cdot)$ is defined in **A2)**, is a uniform upper bound for $\hat{\Phi}_k$. Recall x_0 denotes the initial condition of (4). Proposition 3.3 guarantees an optimal solution exists for all $k \geq 0$ and $\hat{x}_0 \in \mathbb{X}_0$. Assumption **A2)** and condition **C1)** guarantee, for all $T \geq N$,

$$\hat{\Phi}_T - \hat{\Phi}_{T-N} \geq \sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\mathrm{mh}}, \hat{v}_{k|T-1}^{\mathrm{mh}}\right). \qquad (6)$$

We proceed using an induction argument. For $T \leq N$, assumption **A3)**, optimality, and condition **C2)** imply

$$\begin{aligned}\gamma(\|x_0 - \hat{x}_0\|) \geq& \Gamma(x_0) \geq \sum_{k=0}^{T-1} L_k(w_{k|\infty}, v_{k|\infty}) + \Gamma(x_{0|\infty})\\ \geq& \mathcal{Z}_T(x_{T|\infty}) \geq \hat{\mathcal{Z}}_T(x_{T|\infty}).\end{aligned}$$

Condition **C1)** guarantees $\hat{\mathcal{Z}}_T(x_{T|\infty}) \geq \hat{\Phi}_T$ and, therefore, $\gamma(\|x_0 - \hat{x}_0\|) \geq \hat{\Phi}_T$. Let us now assume $\mathcal{Z}_{T-N}(x_{T-N|\infty}) \geq \hat{\mathcal{Z}}_{T-N}(x_{T-N|\infty})$ for the induction argument. Utilizing the optimality principle, we have, for all $T > N$

$$\begin{aligned}&\gamma(\|x_0 - \hat{x}_0\|)\\ &\geq \min_{z,\{w_k\}_{k=T-N}^{T-1}} \left\{ \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \mathcal{Z}_{T-N}(z) : z \in \mathcal{R}_{T-N}, (z,\{w_k\}) \in \Omega_T^N, \ x(T; z, T-N, \{w_j\}) = x_{T|\infty} \right\}\\ &= \mathcal{Z}_T(x_{T|\infty}), \text{(by optimality)}\\ &\geq \min_{z,\{w_k\}_{k=T-N}^{T-1}} \left\{ \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \hat{\mathcal{Z}}_{T-N}(z) : (z,\{w_k\}) \in \Omega_T^N, \ x(T; z, T-N, \{w_j\}) = x_{T|\infty} \right\}\\ &\geq \hat{\mathcal{Z}}_T(x_{T|\infty}). \text{ (by the induction assumption } \mathbf{C2}).\end{aligned}$$

Condition **C1)** guarantees $\hat{\mathcal{Z}}_T(x_{T|\infty}) \geq \hat{\Phi}_T$ for all $T \geq 0$. The sequence $\{\hat{\Phi}_k\}$, therefore, is monotone nondecreasing and bounded above by $\gamma(\|x_0 - \hat{x}_0\|)$. Hence, it is convergent, and the partial sum

$$\sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\mathrm{mh}}, \hat{v}_{k|T-1}^{\mathrm{mh}}\right) \to 0$$

as $T \to \infty$, because the summation in (6) is nonnegative. Lemma 2.5 (with $\mathbb{W}_k = \{0\}$, $\mathbb{V}_k = \{0\}$, and $i = T$) guarantees the estimation error $\|x(T; x_0, 0) - \hat{x}_T^{\mathrm{mh}}\| \to 0$ as claimed.

To prove stability, let $\epsilon > 0$ and choose $\varrho > 0$ as specified by Lemma 2.5 (with $\mathbb{W}_k = \{0\}$, $\mathbb{V}_k = \{0\}$, and $i = T$) such that if

$$\sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\mathrm{mh}}, \hat{v}_{k|T-1}^{\mathrm{mh}}\right) \leq \varrho$$

then $\|x(T; x_0, 0) - \hat{x}_T^{\mathrm{mh}}\| \leq \epsilon$ for all $T \geq N_o$. If we choose $\delta > 0$ such that $\delta < \gamma^{-1}(\varrho)$ (the existence of $\gamma^{-1}(\cdot)$ follows from Fact 2.2), then we obtain the following inequality for all $T \geq N \geq N_o$:

$$\begin{aligned}\gamma(\delta) \geq \Gamma(x_0) \geq& \sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\mathrm{mh}}, \hat{v}_{k|T-1}^{\mathrm{mh}}\right)\\ &+ \hat{\mathcal{Z}}_{T-N}\left(\hat{x}_{T-N|T-1}^{\mathrm{mh}}\right)\\ \geq& \sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\mathrm{mh}}, \hat{v}_{k|T-1}^{\mathrm{mh}}\right).\end{aligned}$$

Hence, if the initial estimation error $\|x_0 - \hat{x}_0\| \leq \delta$, then the estimation error

$$\|x(T; x_0, 0) - \hat{x}_T^{\mathrm{mh}}\| \leq \epsilon$$

for all $T \geq N$ as claimed. □

When the system dynamics are nonlinear, we are unable in general to construct an approximate arrival cost that satisfies condition **C2)** with the exception of the obvious choice $\hat{\mathcal{Z}}_T(\cdot) = \hat{\Phi}_T$. As the proof of Proposition 3.4 demonstrates, condition **C2)** is sufficient to guarantee $\gamma(\|x_0 - \hat{x}_0\|)$ is a uniform upper bound to the optimal cost $\hat{\Phi}_k$ for all $k \geq 0$. While global satisfaction of the inequality (5) in **C2)** is ideal, we may circumvent the issue by explicitly ensuring $\gamma(\cdot)$ is a uniform bound in nominal application. Suppose the sequence of approximate arrival costs $\{\hat{\mathcal{Z}}_j(\cdot)\}_{j=0}^{\infty}$ satisfies condition **C1)**. The purpose of condition **C2)** is to ensure the sequence $\{\hat{\mathcal{Z}}_j(x_{j|\infty})\}$ is monotone nonincreasing [see **A3)**]

$$\begin{aligned}\hat{\mathcal{Z}}_T(x_{T|\infty}) \leq& \min_{z,\{w_k\}_{k=T-N}^{T-1}} \left\{ \sum_{k=T-N}^{T-1} L_k(w_k, v_k) + \hat{\mathcal{Z}}_{T-N}(z) : (z,\{w_k\}) \in \Omega_T^N, \ x(T; z, T-N, \{w_j\}) = x_{T|\infty} \right\}\\ \leq& \sum_{k=T-N}^{T-1} L_k(w_{k|\infty}, v_{k|\infty}) + \hat{\mathcal{Z}}_{T-N}(x_{T-N|\infty}).\end{aligned} \qquad (7)$$

Rather than rely on the general structure of the sequence $\{\hat{\mathcal{Z}}_j(\cdot)\}$ to satisfy the inequality (7), we may force the sequence $\{\hat{\mathcal{Z}}_j(x_{j|\infty})\}$ to be monotone nonincreasing explicitly by scaling the approximate arrival costs

$$\hat{\mathcal{Z}}_j(\cdot) \leftarrow \beta_j\left(\hat{\mathcal{Z}}_j(\cdot) - \hat{\Phi}_j\right) + \hat{\Phi}_j$$

where $\beta_j \in [0, 1]$.

If we knew the sequence $\{x_{k|\infty}\}_{k=0}^{\infty}$ defined in **A3)**, then enforcing the inequality (7) is easy. It is sufficient to scale $\hat{\mathcal{Z}}_T(\cdot)$ such that the inequality (7) is satisfied. The problem is that we rarely know of a sequence satisfying **A3)** *a priori* without first solving a full information estimation problem. However, to satisfy the inequality (7) at time T, we need only to know the last N elements of the sequence $\{x_{k|\infty}, w_{k|\infty}\}_{k=T-N}^{T-1}$. Even this information is unavailable *a priori*, though we may obtain it online. What we need to generate online is a feasible state sequence $\{x_k^0, w_k^0\}_{k=T-N}^{T-1}$ that is bounded by the initial estimation error in nominal application. We can generate this feasible sequence using $\hat{\mathcal{Z}}_{T-N}(\cdot) = \hat{\Phi}_{T-N}$. Recall from Remark 3.1 that this choice for the approximate arrival cost yields a stable constrained observer. Once we have a feasible sequence, we can scale $\hat{\mathcal{Z}}_T(\cdot)$ such that it satisfies (7).

Consider the MHE problem where we choose $\hat{\mathcal{Z}}_T(\cdot) = \hat{\Phi}_T$. We formulate this estimation problem as the following optimal control problem[2]

$$P_3(T): \quad \psi_T^* = \min_{z,\{w_k\}_{k=T-N}^{T-1}} \left\{ \sum_{k=T-N}^{N-1} L_k(w_k, v_k) \right. \left. : (z, \{w_k\}) \in \Omega_T^N \right\}.$$

For $T \leq N$, $P_3(T)$ is defined to be $P_1(T)$. The solution to $P_3(T)$ is the pair

$$\left(z^*, \{\hat{w}_{k|T-1}^0\}_{k=T-N}^{T-1}\right)$$

and that optimal pair yields an estimate $\{\hat{x}_{k|T-1}^0\}_{k=T-N}^{T}$ of the actual sequence $\{x_k\}$, where

$$\hat{x}_{k|T-1}^0 := x\left(k, z^*, T-N, \left\{\hat{w}_{j|T-1}^0\right\}\right).$$

It follows that $\hat{x}_0^0 = \hat{x}_0$. We formulate the estimation strategy as the following algorithm.

`Estimation Algorithm`
`Data` $N \in \mathbb{N}$.
`Initialization:` `For` $T \leq N$ `do:`
`1. Solve` $P_2(T)$ `to generate` $\{\hat{x}_k\}_{k=1}^N$ `and` $\{\hat{\Phi}_k\}_{k=1}^N$.
`2. Solve` $P_3(T)$ `to obtain` $\hat{x}_{0|N-1}^0$ `and` $\{\psi_k^*\}_{k=1}^N$.
`3. For` $k = 1, \ldots, N$, `set` $U_k = \psi_k^* + \Gamma(\hat{x}_{0|N-1}^0)$.
`Step 1` `For` $T > N$ `do:`
`1. Solve` $P_3(T)$ `to obtain` $\hat{x}_{T-N|T-1}^0$ `and` ψ_T^*.
`2. Set` $U_T = \psi_T^* + U_{T-N}$.
`3. Construct` $\hat{\mathcal{Z}}_{T-N}(\cdot)$ `so that it satisfies` **`C1`**.
`4. Set`

$$\beta_{T-N} = \max_{\beta \in [0,1]} \left\{ \beta : \beta\left(\hat{\mathcal{Z}}_{T-N}(\hat{x}_{T-N|T-1}^0) - \hat{\Phi}_{T-N}\right) \right. \left. + \hat{\Phi}_{T-N} \leq U_{T-N} \right\}.$$

`5. Set`

$$\hat{\mathcal{Z}}_{T-N}(\cdot) \leftarrow \beta_{T-N}\left(\hat{\mathcal{Z}}_{T-N}(\cdot) - \hat{\Phi}_{T-N}\right) + \hat{\Phi}_{T-N}$$

`6. Solve` $P_2(T)$ `and obtain` $\hat{x}_T$ `and` $\hat{\Phi}_T$.
`Step 2` `Let` $T \leftarrow T+1$. `Go to Step 1.`

We have constructed stable variants of the proposed estimation algorithm including suboptimal algorithms, where global solutions to the associated optimal control problem are not necessary. The interested reader is directed to [4].

Remark 3.5: If we choose

$$\hat{\mathcal{Z}}_j(x) = (x - \hat{x}_j)^T \Pi^{-1} (x - \hat{x}_j) + \hat{\Phi}_j$$

where the matrix Π^{-1} is symmetric positive semi-definite, then **C1)** is automatically satisfied; let $\bar{\gamma}(\cdot) = (1 + \lambda_{\max}(\Pi^{-1}))\|(\cdot)\|^2$.

The stability of the proposed algorithm relies on the stability of the estimator defined by $P_3(T)$. We know from Proposition 3.4 that $\|\hat{x}_T^0 - x(T; x_0, 0)\| \to 0$ as $T \to \infty$. More importantly, we know that the sequence $\{U_k\}$ is bounded.

Proposition 3.6: If assumptions **A0)–A3)** hold, the system (1) is uniformly observable, and $N \geq N_o$, then, for all $\hat{x}_0 \in \mathbb{X}_0$, MHE using the estimation algorithm is an asymptotically stable observer for (4).

Proof: From the preceding arguments (see the proof of Proposition 3.4), it suffices to show U_T is bounded uniformly for all $k \geq 0$ by $\|x_0 - \hat{x}_0\|$. Let $V = \gamma(\|x_0 - \hat{x}_0\|) + \Gamma(\hat{x}_{0|N-1}^0)$. Optimality guarantees $\sum_{k=T-N}^{T-1} L_k(w_{k|\infty}, v_{k|\infty}) \geq \psi_T^*$ for all $T \geq N$. Hence, by **A3)**, we have $U_k \leq V$ for all $k \geq N$. By construction, for $T \geq N$, $\hat{\mathcal{Z}}_{T-N}(\hat{x}_{T-N|T-1}^0) \leq U_{T-N}$. Because $\left(\hat{x}_{T-N|T-1}^0, \{\hat{w}_{k|T-1}^0\}\right) \in \Omega_T^N$, optimality implies $\hat{\Phi}_T \leq U_T$. Hence, the sequence $\{\hat{\Phi}_j\}$ is bounded above by V and, consequently, $\|x(T; x_0, 0) - \hat{x}_T\| \to 0$ as $T \to \infty$.

We now establish that V is bounded by $\|x_0 - \hat{x}_0\|$. By assumption **A3)**, $\psi_N^* \leq \gamma(\|x_0 - \hat{x}_0\|)$ and, by Lemma 2.5 (with $\mathbb{W}_k = \{0\}$, $\mathbb{V}_k = \{0\}$, and $i = T - N$)

$$\begin{aligned}\|\hat{x}_{0|N-1}^0 - x_0\| &\leq \theta(\psi_N^*)\\ &\leq \theta\left(\gamma(\|x_0 - \hat{x}_0\|)\right).\end{aligned}$$

Hence, we obtain

$$\begin{aligned}V &\leq \gamma(\|x_0 - \hat{x}_0\|) + \bar{\gamma}\left(\left\|\hat{x}_{0|N-1}^0 - \hat{x}_0\right\|\right)\\ &\leq \gamma(\|x_0 - \hat{x}_0\|) + \bar{\gamma}\left(\|\hat{x}_{0|N-1}^0 - x_0\| + \|x_0 - \hat{x}_0\|\right)\\ &\leq \gamma(\|x_0 - \hat{x}_0\|) + \bar{\gamma}\left(\theta\left(\gamma(\|x_0 - \hat{x}_0\|)\right) + \|x_0 - \hat{x}_0\|\right)\\ &:= \omega(\|x_0 - \hat{x}_0\|)\end{aligned}$$

where $\bar{\gamma}(\cdot)$ results from applying condition **C1)** and $\omega(\cdot)$ is a K-function. The existence of the K-function $\omega(\cdot)$ follows from Fact 2.3. □

We desire $\beta_T = 1$ when $\hat{\mathcal{Z}}_T(\cdot)$ satisfies condition **C2)**. If we assume $x(k; x_0, 0) \in \mathbb{X}_k$ for all $k \geq 0$, then optimality and the observability assumption imply $\hat{x}_{T-N|T-1}^0 = x(T-N; x_0, 0)$ for all $T \geq N$ and, as a consequence, $U_T = 2\Gamma(x_0)$. It follows by optimality and condition **C2)** that for $T \geq N$

$$\begin{aligned}\hat{\mathcal{Z}}_T(\hat{x}_{T-N|T-1}^0) &= \hat{\mathcal{Z}}_T(x(T-N; x_0, 0))\\ &\leq 2\Gamma(x_0).\end{aligned}$$

[2]Adding a constant to the objective function does not affect the answer. For simplicity, we choose $\hat{\mathcal{Z}}_T(\cdot) = 0$

Therefore, $\beta_T = 1$. When the constraints only satisfy **A3)** or when we consider suboptimal algorithms, then the estimation algorithm does not guarantee $\beta_T = 1$ when the sequence $\{\hat{\mathcal{Z}}_T^0(\cdot)\}$ satisfies condition **C2)**. To guarantee $\beta_T = 1$, one can modify the estimation algorithm (see [4]).

A. Bounded Disturbances

In this section, we investigate the stability properties of MHE when the sets $\mathbb{X}_0$, $\mathbb{W}_k$, and $\mathbb{V}_k$ are uniformly compact. We demonstrate under these condition that the estimation error is bounded. The bounds that we derive are conservative and not constructive, though they illustrate the performance of MHE in the presence of noise. Our arguments build on many of the results discussed in Sections I and II; for brevity, we freely make use of those results.

D0) There exists scalars $\delta_w > 0$, $\delta_v > 0$, and $\delta_x > 0$ such that $\mathbb{W}_k \subset B_{\delta_w}$, $\mathbb{V}_k \subset B_{\delta_v}$, $\mathbb{X}_0 \subset B_{\delta_x}$ for all $k \geq 0$.

Throughout this section, we denote the dynamics of the true system by the sequences $\{x_k\}$, $\{w_k\}$, and $\{v_k\}$, where by assumption $x_0 \in \mathbb{X}_0$, $x_k \in \mathbb{X}_k$, $w_k \in \mathbb{W}_k$, and $v_k \in \mathbb{V}_k$ for all $k \geq 0$. In other words, the dynamics of the true system obey the constraints.

Proposition 3.7: If the assumptions **A0)–A3)**, the sequence $\{\hat{\mathcal{Z}}_j(\cdot)\}$ satisfies conditions **C1)** and **C2)**, the system (1) is uniformly observable, $N \geq N_o$, and the constraints satisfy condition **D0)**, then the estimation error $\|x_T - \hat{x}_T^{\rm mh}\|$ for MHE is bounded for all $T \geq N$.

Proof: We assume throughout the proof that $T \geq N$. Proposition 3.3 guarantees a solution exists for all $k \geq 0$. Let $b := \gamma(\delta_w + \delta_v)$. Assumption **A2)** and condition **C1)** guarantee, for all $T \geq N$, that

$$\hat{\Phi}_T - \hat{\Phi}_{T-N} \geq \sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\rm mh}, \hat{v}_{k|T-1}^{\rm mh}\right).$$

From the proof of Proposition 3.4, we know under the stated assumptions that

$$\mathcal{Z}(x_T) \geq \hat{\mathcal{Z}}(x_T)$$

for all $T \geq 0$. Optimality then implies

$$Tb + \gamma(\delta_x) \geq \hat{\Phi}_T$$

as optimality implies $\hat{\Phi}_T \leq \hat{\mathcal{Z}}(x_T)$. As $\hat{x}_{T-N}^{\rm mh}$ is feasible for problem $P_2(T)$ for all sequences $\{w_k, v_k\}_{k=T-N}^{T-1} \in \mathbb{W}^N \times \mathbb{V}^N$, optimality implies

$$\begin{aligned} Nb \geq& \hat{\Phi}_T - \hat{\Phi}_{T-N} \\ \geq& \sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\rm mh}, \hat{v}_{k|T-1}^{\rm mh}\right). \end{aligned}$$

Lemma 2.5, consequently, states that the estimation error is bounded as claimed. □

Corollary 3.8: If the assumptions **A0)–A3)**, system (1) is uniformly observable, $N \geq N_o$, and the constraints satisfy condition **D0)**, then the estimation error $\|x_T - \hat{x}_T^{\rm mh}\|$ for the estimation algorithm is bounded for all $T \geq N$.

Proof: By construction $U_T - U_{T-N} \leq Nb$ and $U_T - \hat{\Phi}_T < \Gamma(\hat{x}_{0|N-1}^0)$ where $b := \gamma(\delta_w + \delta_v)$. Hence

$$\begin{aligned} Nb + \Gamma(\hat{x}_{0|N-1}^0) \geq& \hat{\Phi}_T - \hat{\Phi}_{T-N} \\ \geq& \sum_{k=T-N}^{T-1} L_k\left(\hat{w}_{k|T-1}^{\rm mh}, \hat{v}_{k|T-1}^{\rm mh}\right). \end{aligned}$$

Lemma 2.5 , consequently, states that the estimation error is bounded as claimed. □

IV. Examples of Inequality Constraints Yielding Improved Estimates

In this section, we demonstrate how inequality constraints improve the state estimate when the disturbances are bounded. We first consider a linear example where we use the Kalman filter, the unconstrained full information estimator, as a benchmark. We then consider a nonlinear example and use the extended Kalman filter (EKF) as a benchmark.

Consider the following discrete-time system

$$\begin{aligned} x_{k+1} =& \begin{bmatrix} 0.99 & 0.2 \\ -0.1 & 0.3 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_k \\ y_k =& \begin{bmatrix} 1 & -3 \end{bmatrix} x_k + v_k. \end{aligned} \tag{8}$$

We assume $\{v_k\}$ is sequence of independent, zero mean, normally distributed random variables with covariance 0.01, and $w_k = |z_k|$ where $\{z_k\}$ is a sequence of independent, zero mean, normally distributed random variables with unit covariance. We also assume the initial state x_0 is normally distributed with zero mean and covariance equal to the identity.

We formulate the constrained estimation problem with $Q = 1$, $R = 0.01$, $\Pi_0 = 1$, and $\hat{x}_0 = 0$. For the MHE, we choose $N = 10$ and

$$\hat{\mathcal{Z}}_\tau(z) = (z - \hat{x}_\tau^{mh})^T \Pi_\tau^{-1} (z - \hat{x}_\tau^{mh}) + \hat{\Phi}_\tau$$

where the sequence $\{\Pi_j\}$ is obtained by solving the matrix Riccati (3). As stated previous, this choice for the approximate arrival cost satisfies condition **C2)**. To capture our knowledge of the random sequence w_k, we add the inequality constraint $w_k \geq 0$. Note, this formulation yields the *optimal* Bayesian estimate. A comparison of the Kalman filter and MHE is shown in Fig. 1. For a benchmark, we used the sum squre estimation error

$$\sum_{k=0}^{T} \left(x_k^{(j)} - \hat{x}_k^{(j)}\right)^2$$

where $x^{(j)}$ denotes the jth entry of the vector x. For $x^{(1)}$, the average sum square estimation error based on 100 trials was 1194.45 for the Kalman filter and 36.08 for MHE. For $x^{(2)}$, the average square estimation error was 131.15 for the Kalman filter and 81.60 for MHE. As expected, the performance of the constrained estimators is superior to the Kalman filter, because the constrained estimators possess, with the addition of the inequality constraints, the proper statistics of the disturbance sequence w_k. Hence, the constrained estimation problem formulated above accurately models the random variable w_k.

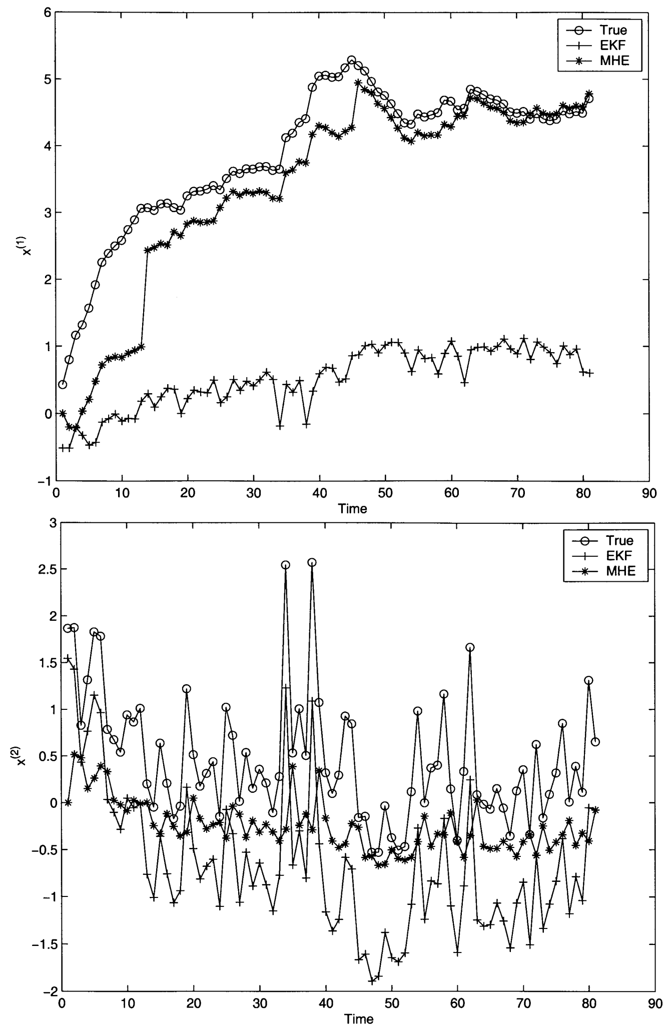


Fig. 1. Comparison of estimators for model (8).

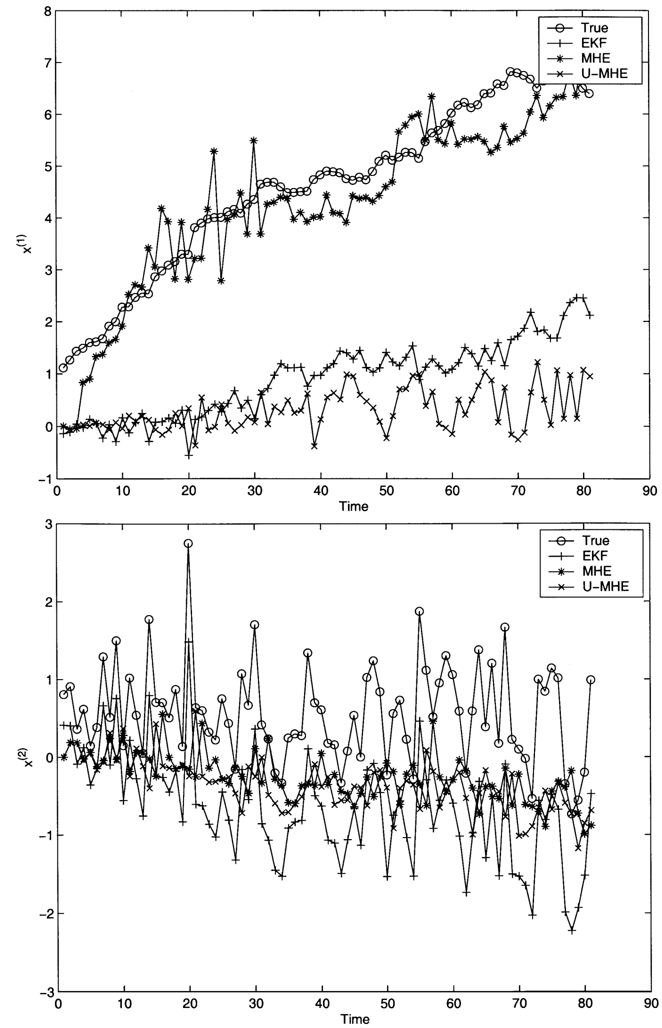


Fig. 2. Comparison of estimators for model (9).

To compare the performance of MHE and EKF, we consider the following nonlinear perturbation of the model (8):

$$x_{k+1}^{(1)} = 0.99x_k^{(1)} + 0.2x_k^{(2)} \tag{9a}$$

$$x_{k+1}^{(2)} = -0.1x_k^{(1)} + \frac{0.5x_k^{(2)}}{1+(x^{(2)})^2} + w_k \tag{9b}$$

$$y_k = x_k^{(1)} - 3x_k^{(1)} + v_k. \tag{9c}$$

The disturbances are modeled as random variables with the same distributions as the previous example. We formulate MHE as above with the exception that the sequence $\{\Pi_j\}$ is obtained using an extended Kalman filter update. A comparison of the EKF, unconstrained MHE (U-MHE), and MHE is shown in Fig. 2. Once again, the unconstrained EKF estimate diverges while the MHE estimate is able to track the state. If we compare constrained versus unconstrained MHE, then the conclusions are similar: constraints are necessary for an accurate state estimate. For $x^{(1)}$, the average sum square estimation error based on 100 trials was 888.97 for the EKF and 66.58 for MHE. For $x^{(2)}$, the average square estimation error was 97.66 for the EKF and 76.83 for MHE.

We repeated both examples using bounded noise, where $w_k = \min\{|z_k|, 2\}$. The results were similar. For the linear example, the average sum square estimation error for the Kalman filter based on 100 trials was 1149.10 for $x^{(1)}$ and 126.04 for $x^{(2)}$. The average sum square estimation error for MHE based on 100 trials was 37.44 for $x^{(1)}$ and 74.94 for $x^{(2)}$. For the nonlinear example, the average sum square estimation error for the extended Kalman filter based on 100 trials was 852.25 for $x^{(1)}$ and 93.63 for $x^{(2)}$. The average sum square estimation error for MHE based on 100 trials was 50.07 for $x^{(1)}$ and 69.99 for $x^{(2)}$.

At each time step, the solution of the quadratic program took approximately a tenth of a second and the solution of the nonlinear program took approximately 3 s on a desktop computer. A single realization involve 80 data points took approximately 10 s for the linear example and 3 min for the nonlinear examples. The time required for either the Kalman filter or the extended Kalman filter was negligible. All computations were performed in GNU Octave on a 500-MHz processor. No effort was made to improve the efficiency of either computation.

V. Conclusion

In this paper, we investigated MHE as an online optimization strategy for estimating the state of constrained discrete-time systems. The practical significance of MHE is the ability to incorporate constraints explicitly. This feature distinguishes MHE from other strategies such as extended Kalman filtering and output error linearization. Furthermore, if the estimation problem translates into a problem of the form $P_1(T)$, then we believe MHE is a natural engineering approximation to the full information problem, because the structure of MHE is not dictated by stability, but rather by performance and practicality. Stability results if one judiciously approximates the past data.

One limitation of MHE is the need for global solutions to the optimization problems $P_2(T)$ and $P_3(T)$. This computational requirement presents a barrier to online implementation. Aside from the computational burden, optimization may not yield global solutions unless the problem is convex. Strategies exist for finding a global solution, though they are currently impractical for online implementation. The difficulty in global optimization is not finding a solution, but rather verifying whether a particular solution is global. Unless global information such as lower bounds or Lipschitz constants are available, one needs to sample a dense subset of the decision space in order to guarantee a particular solution is global [41]. In results not discussed here [4], we propose a stable suboptimal version of MHE that does not require a global solution. This algorithm is similar to the suboptimal version of receding horizon control first proposed by Michalska and Mayne [42] and further developed in discrete time by Scokaert, Mayne, and Rawlings [43].

The strength and weakness of MHE is the use of constrained optimization. For many systems, the optimization problems can be solved in a few seconds on a desktop computer using standard software such as Matlab. However, for some estimation problems, MHE is too slow. With the increasing power of computers and improved algorithms (i.e algorithms now solve quadratic programs in polynomial time), MHE will become an alternative for an expanding class of constrained state estimation problems in the near future.

APPENDIX I

A. Proof of Lemma 2.5

Proof: Recall x_k denotes the true state of (1). We now make make the following definitions:

$$\begin{aligned}\bar{x}_k &:= x(k;\hat{x}_{T-N|T-1},T-N)\\ \bar{y}_k &:= y(k;\hat{x}_{T-N|T-1},T-N)\\ \tilde{x}_k &:= x(k;x_{T-N},T-N)\\ \tilde{y}_k &:= y(k;x_{T-N},T-N)\\ \hat{v}_{k|T-1} &:= h_k(x_k)+v_k-h_k(\hat{x}_{k|T-N})\\ &= y_k - h_k(\hat{x}_{k|T-N}).\end{aligned}$$

Employing the triangle inequality, we obtain the bound

$$\|x_i-\hat{x}_i\| \le \|\hat{x}_i-\bar{x}_i\|+\|\bar{x}_i-\tilde{x}_i\|+\|x_i-\tilde{x}_i\|. \qquad (10)$$

By the Lipschitz continuity of $f_k(\cdot)$, we have the inequalities

$$\|\hat{x}_i-\bar{x}_i\| \le \sum_{k=T-N}^{i-1} c_f^{i-k}\|\hat{w}_{k|T-1}\| \qquad (11a)$$

$$\|x_i-\tilde{x}_i\| \le \sum_{k=T-N}^{i-1} c_f^{i-k}\|w_k\| \qquad (11b)$$

for all $i \in \{T-N, T-N+1, \ldots, T\}$. Likewise, assumption **A2)** implies

$$\|\hat{w}_{k|T-1}\| \le \eta^{-1}\left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1},\hat{v}_{k|T-1})\right) \qquad (12a)$$

$$\|\hat{v}_{k|T-1}\| \le \eta^{-1}\left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1},\hat{v}_{k|T-1})\right) \qquad (12b)$$

where the existence of K-function $\eta^{-1}(\cdot)$ follows from Fact 2.2. Hence

$$\|\hat{x}_i-\bar{x}_i\| \le \sum_{k=T-N}^{i-1} c_f^{i-k}\eta^{-1} \times \left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1},\hat{v}_{k|T-1})\right).$$

In the remaining steps of the proof, we demonstrate that the quantity $\|\bar{x}_i-\tilde{x}_i\|$ is bounded.

By definition

$$\sum_{k=T-N}^{T-1}\|\hat{v}_k\| = \sum_{k=T-N}^{T-1}\|h_k(x_k)-h_k(\hat{x}_k)+v_k\|.$$

By repeated application of the inverse triangle inequality and utilizing the observability condition, we obtain the inequality

$$\begin{aligned}&\sum_{k=T-N}^{T-1}\|\hat{v}_k\|+\|v_k\| + \|h_k(\hat{x}_k)-\bar{y}_k\|+\|h_k(x_k)-\tilde{y}_k\|\\ &\quad\ge \sum_{k=T-N}^{T-1}\|\bar{y}_k-\tilde{y}_k\|\\ &\quad\ge \varphi(\|\bar{x}_{T-N}-\tilde{x}_{T-N}\|).\end{aligned}$$

By the Lipschitz continuity of $f_k(\cdot)$, we obtain the inequality

$$\|\bar{x}_i-\tilde{x}_i\| \le (1+c_f)^i\varphi^{-1}\left(\sum_{k=T-N}^{T-1}\|\hat{v}_k\|+\|v_k\| + \|h_k(\hat{x}_k)-\bar{y}_k\|+\|h_k(x_k)-\tilde{y}_k\|\right) \qquad (13)$$

where the existence of the K-function $\varphi^{-1}(\cdot)$ follows from Fact 2.2. By the Lipschitz continuity of $h_k(\cdot)$, (11a), and (12a), we obtain the inequality

$$\begin{aligned}\sum_{k=T-N}^{T-1}\|h_k(\hat{x}_k)-\bar{y}_k\| &\le \sum_{j=T-N}^{T-1}\sum_{k=T-N}^{j} c_h c_f^{j-k}\|\hat{w}_{k|T-1}\|\\ &\le \sum_{j=T-N}^{T-1}\sum_{k=T-N}^{j} c_h c_f^{j-k}\eta^{-1} \times\left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1},\hat{v}_{k|T-1})\right).\end{aligned}$$

Likewise, By the Lipschitz continuity of $h_k(\cdot)$, (11b), and (12b), we obtain the inequality

$$\sum_{k=T-N}^{T-1} \|h_k(x_k) - \tilde{y}_k\| \leq \sum_{j=T-N}^{T-1} \sum_{k=T-N}^{j} c_h c_f^{j-k} \|w_k\|.$$

Substituting into (13), we obtain the inequality

$$\begin{aligned} \|\bar{x}_i - \tilde{x}_i\| \leq &(1+c_f)^i \varphi^{-1} \\ &\times \left(N\eta^{-1} \left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1}, \hat{v}_{k|T-1}) \right) \right. \\ &+ N\delta_v + \sum_{j=T-N}^{T-1} \sum_{k=T-N}^{j} c_h c_f^{j-k} \\ &\times \left(\eta^{-1} \left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1}, \hat{v}_{k|T-1}) \right) \right. \\ &\left. \left. + \delta_w \right) \right). \end{aligned}$$

Substituting the aforementioned expressions in (10), we obtain the inequality

$$\begin{aligned} \|x_i - \hat{x}_i\| \leq &(1+c_f)^i \varphi^{-1} \\ &\times \left(N\eta^{-1} \left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1}, \hat{v}_{k|T-1}) \right) \right. \\ &+ N\delta_v + \sum_{j=T-N}^{T-1} \sum_{k=T-N}^{j} c_h c_f^{j-k} \\ &\times \left(\eta^{-1} \left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1}, \hat{v}_{k|T-1}) \right) \right. \\ &+\delta_w)) \\ &+ \sum_{k=T-N}^{T-1} c_f^{T-k} \left(\eta^{-1} \right. \\ &\left. \times \left(\sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1}, \hat{v}_{k|T-1}) \right) + \delta_w \right). \end{aligned}$$

Collectively defining the terms on the right hand side of the inequality as a function $\delta(\cdot)$, we obtain the following bound of the estimation error:

$$\begin{aligned} &\|\bar{x}_i - \tilde{x}_i\| \\ &\leq \theta_{\sigma_w, \sigma_v} \left(\left\| \left(\left\{ \sum_{k=T-N}^{T-1} L_k(\hat{w}_{k|T-1}, \hat{v}_{k|T-1}) \right\}, \delta_w, \delta_v \right) \right\| \right). \end{aligned}$$

Facts 2.2 and 2.3 guarantee $\theta(\cdot)$ is K-function as it is a positive linear combination and composition of the K-functions $\eta^{-1}(\cdot)$ and φ^{-1}. □

APPENDIX II

A. Proof of Proposition 3.3

Proof: For $T \leq N$, existence is established by routine application of the Weierstrass Maximum Theorem (see [4] for the specific details). Now consider $T > N$ and let

$$\hat{\Phi}_T^1 = \sum_{k=T-N}^{T-1} L_k(w_{k|\infty}, v_{k|\infty}) + \hat{\mathcal{Z}}_{T-N}(x_{T-N|\infty})$$

denote the finite cost, by assumption **A2)** and property **C1)**, associated with the feasible sequence $x_{T-N|\infty}$ and $\{w_{k|\infty}\}_{k=T-N}^{T-1}$ specified in assumption **A3)**. Consider the set

$$\begin{aligned} \Lambda = \Big\{ &z, \{w_k\}_{k=T-N}^{T-1} \\ &: (z, \{w_k\}) \in \Omega_T^N, \hat{\Phi}_T(z, \{w_k\}) \leq \hat{\Phi}_T^1 \Big\}. \end{aligned}$$

A solution exists under the stated assumption by application of the Weierstrass Maximum Theorem if the set Λ is bounded. Assumption **A2)** guarantees the sequence $\{w_k, v_k\}_{k=T-N}^{T-1}$ is bounded: $\|w_k\| + \|v_k\| \leq 2\eta^{-1}(\hat{\Phi}_T^1)$. We conclude by demonstrating z is bounded. If we employ the inverse triangle inequality, we obtain

$$\begin{aligned} \sum_{k=T-N}^{T-1} \|v_k\| &= \sum_{T-N}^{T-1} \|y_k - \bar{y}_k\| \\ &\geq \sum_{k=T-N}^{T-1} \|\bar{y}_k - y_{k|\infty}\| - \|y_k - y_{k|\infty}\| \end{aligned}$$

where $\bar{y}_k := y(k; z, T-N, \{w_j\})$ and $y_{k|\infty} := y(k; x_{T-N|\infty}, T-N, \{w_{,|\infty}\})$. Rearranging the inequality, we obtain

$$\begin{aligned} \sum_{k=T-N}^{T-1} \|v_k\| + \|y_k - y_{k|\infty}\| &= \sum_{k=T-N}^{T-1} \|v_k\| + \|v_{k|\infty}\| \\ &\geq \sum_{k=T-N}^{T-1} \|\bar{y}_k - y_{k|\infty}\|. \end{aligned}$$

If we employ again the inverse triangle inequality, we obtain

$$\begin{aligned} &\sum_{k=T-N}^{T-1} \|\bar{y}_k - y_{k|\infty}\| \\ &\quad \geq \sum_{k=T-N}^{T-1} \|\bar{y}_k - y(k; x_{T-N|\infty}, T-N)\| \\ &\qquad - \|y(k; x_{T-n|\infty}, T-N) - y_{k|\infty}\| \\ &\quad \geq \sum_{k=T-N}^{T-1} \|y(k; z, T-N) - y(k; x_{T-N|\infty}, T-N)\| \\ &\qquad - \left(\sum_{k=T=N}^{T-1} \|\bar{y}_k - y(k; z, T-N)\| \right. \\ &\qquad \left. + \|y_{k|\infty} - y(k; x_{T-N|\infty}, T-N)\| \right). \end{aligned}$$

Rearranging the inequality and applying the observability assumption, we obtain the inequality

$$\sum_{T-N}^{T-1} \|y_{k|\infty} - \bar{y}_k\| + \|\bar{y}_k - y(k; z, T-N)\| + \|y_{k|\infty} - y(k; x_{T-N|\infty}, T-N)\| \geq \varphi(\|x_{T-N|\infty} - z\|).$$

The first quantity $\|y_{k|\infty} - \bar{y}_k\|$ is bounded, using the triangle inequality, by $\|v_k\| + \|v_{k|\infty}\|$ and, consequently, by $2N\eta^{-1}(\hat{\Phi}^1)$. To show the last two quantities are bounded, we employ assumption **A0)** to obtain the following inequality:

$$\begin{aligned}\|\bar{y}_k - y(k; z, T-N)\| &\leq c_h \sum_{j=T-N}^{k-1} c_f^{k-j} \|w_k\| \\ &\leq c_h \sum_{j=T-N}^{k-1} c_f^{k-j} 2N\eta^{-1}(\hat{\Phi}_T^1).\end{aligned}$$

Likewise, we have the inequality

$$\|y_{k|\infty} - y(k; x_{T-N|\infty}, T-N)\| \leq c_h \sum_{j=T-N}^{k-1} c_f^{k-j} \eta^{-1}\left(\hat{\Phi}_T^1\right).$$

Consequently, the quantity $\|x_{T-N|\infty} - z\|$ is bounded, and existence follows as claimed. □

REFERENCES

[1] H. Cox, "On the estimation of state variables and parameters for noisy dynamic systems," *IEEE Trans. Automat. Contr.*, vol. 9, pp. 5–12, Jan. 1964.

[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.

[3] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[4] C. V. Rao, "Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems," Ph.D. dissertaion, Univ. Wisconsin-Madison, Madison, WI, 2000.

[5] T. Başar and P. Bernhard, *H^∞-Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Boston, MA: Birkauser, 1995.

[6] S. Verdu and H. V. Poor, "Abstract dynamic programming models under commutativity conditions," *SIAM J. Control Optim.*, vol. 25, pp. 990–1006, 1987.

[7] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena, 1995, vol. 1.

[8] C. Striebel, "Sufficient statistics in the optimum control of stochastic systems," *J. Math. Anal. Appl.*, vol. 12, pp. 576–592, 1965.

[9] S.-S. Jang, B. Joseph, and H. Mukai, "Comparison of two approaches to on-line parameter and state estimation of nonlinear systems," *Ind. Eng. Chem. Proc. Des. Dev.*, vol. 25, pp. 809–814, 1986.

[10] Y. A. Thomas, "Linear quadratic optimal estimation and control with receding horizon," *Electron. Lett.*, vol. 11, pp. 19–21, Jan. 1975.

[11] W. H. Kwon, A. M. Bruckstein, and T. Kailath, "Stabilizing state-feedback design via the moving horizon method," *Int. J. Control*, vol. 37, no. 3, pp. 631–643, 1983.

[12] B. W. Bequette, "Nonlinear predictive control using multi-rate sampling," *Can. J. Ghem. Eng.*, vol. 69, pp. 136–143, Feb. 1991.

[13] Y. Ramamurthi, P. Sistu, and B. Bequette, "Control-relevant dynamic data reconciliation and parameter estimation," *Comput. Chem. Eng.*, vol. 17, no. 1, pp. 41–59, 1993.

[14] I. Kim, M. Liebman, and T. Edgar, "A sequential error-in-variables method for nonlinear dynamic systems," *Comput. Chem. Eng.*, vol. 15, no. 9, pp. 663–670, 1991.

[15] M. Liebman, T. Edgar, and L. Lasdon, "Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques," *Comput. Chem. Eng.*, vol. 16, no. 10/11, pp. 963–986, 1992.

[16] I. B. Tjoa and L. T. Biegler, "Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems," *Comput. Chem. Eng.*, vol. 15, no. 10, pp. 679–690, 1991.

[17] J. Albuquerque and L. I. Biegler, "Data reconciliation and gross-error detection for dynamic systems," *AIChE J.*, vol. 42, no. 10, pp. 2841–2856, 1996.

[18] J. Albuquerque and L. T. Biegler, "Decomposition algorithms for on-line estimaition with nonlinear DAE models," *Comput. Chem. Eng.*, vol. 21, no. 3, pp. 283–299, 1997.

[19] A. M'hamdi, A. Helbig, O. Abel, and W. Marquardt, "Newton-type receding horizon control and state estimation," in *Proc. 1996 IFAC World Cong.*, San Francisco, CA, 1996, pp. 121–126.

[20] T. Binder, L. Blank, W. Dahmen, and W. Marquardt, "Toward multiscale dynamic data reconciliation," in *NATO ASI on Nonlinear Model Based Process Control*. Norwell, MA: Kluwer, 1998.

[21] A. Bemporad, D. Mignone, and M. Moran, "Moving horizon estimation for hybrid systems and fault detection," presented at the 1999 Amer. Control Conf., San Diego, CA, 1999.

[22] R. Gesthuisen and S. Engell, "Determination of the mass transport in the poly condensation of polyethyleneterephthalate by nonlinear estimation techniques," presented at the 1998 IFAC DYCOPS Symp., Corfu, Greece, 1998.

[23] L. P. Russo and R. E. Young, "Moving horizon state estimation applied to an indus trial polymerization process," presented at the 1999 Amer. Control Conf., San Diego, CA, 1999.

[24] D. G. Robertson and J. H. Lee, "A least squares formulation for state estimation," *J. Proc. Control*, vol. 5, no. 4, pp. 291–299, 1995.

[25] D. G. Robertson, J. H. Lee, and J. B. Rawlings, "A moving horizon-based approach for least-squares state estimation," *AIChE J.*, vol. 42, pp. 2209–2224, Aug. 1996.

[26] D. G. Robertson and J. H. Lee, "On the use of constraints in least squares estimation and control," *Automatica*, 2002.

[27] K. R. Muske, J. B. Rawlings, and J. H. Lee, "Receding horizon recursive state estimation," in *Proc. 1993 Amer. Control Conf.*, June 1993, pp. 900–904.

[28] K. R. Muske and J. B. Rawlings, "Nonlinear moving horizon state estimation," in *Methods of Model Based Process Control*. ser. Nato Advanced Study Institute series: E Applied Sciences 293, R. Berber, Ed. Dordrecht, The Netherlands: Kluwer, 1995, pp. 349–365.

[29] M. L. Tyler and M. Moran, "Stability of constrained moving horizon estimation schemes," in *Preprint A UT96-18*. Zurich, Switzerland: Automatic Control Laboratory, Swiss Federal Inst. Technol., 1996.

[30] M. L. Tyler, "Performance monitoring and fault detection in control systems," Ph.D. dissertation, California Inst. Technol., Pasadena, CA, 1997.

[31] K. V. Ling and K. W. Lim, "Receding horizon recursive state estimation," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1750–1753, Sept. 2000.

[32] W. H. Kwon, P. S. Kim, and P. Park, "A receding horizon Kalman FIR filter for linear continuous-time systems," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 2115–2120, Nov. 1999.

[33] ——, "A receding horizon Kalman FIR filter for discrete time-invariant systems," *IEEE Trans Automat. Contr.*, vol. 44, pp. 1787–1791, Mar. 1995.

[34] G. Zimmer, "State Observation by on-line minimization," *Int. J. Control*, vol. 60, no. 4, pp. 595–606, 1994.

[35] P. E. Moraal and J. W. Grizzle, "Observer design for nonlinear systems with discrete-time measurements," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 395–404, Mar. 1995.

[36] H. Michalska and D. Q. Mayne, "Moving horizon observers and observer-based control," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 995–1006, June 1995.

[37] T. L. Vincent and P. P. Khargonekar, "A class of nonlinear filtering problems arising from drifting sensor gains," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 509–520, Mar. 1999.

[38] A. Alessandri, M. Baglieto, T. Parisini, and R. Zoppoli, "A neural state estimator with bounded errors for nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 2028–2042, Nov. 1999.

[39] H. L. Royden, *Real Analysis*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1988.

[40] S. S. Keerthi and E. G. Gilbert, "Optimal infinite-horizon feedback laws for a general class of constrained discrete-time systems: Stability and moving-horizon approximations," *J. Optim. Theory Appl.*, vol. 57, pp. 265–293, May 1988.

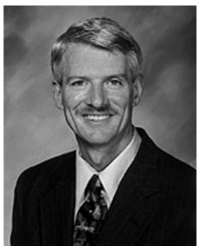
[41] C. P. Stephens and W. Baritompa, "Global optimization requires global information," *J. Optim. Theory Appl.*, vol. 96, pp. 575–588, 1998.

[42] H. Michalska and D. Q. Mayne, "Robust receding horizon control of constrained nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 38, pp. 1623–1633, Nov. 1993.

[43] P. O. M. Scokaert, D. Q. Mayne, and J. B. Rawlings, "Suboptimal model predictive control (feasibility implies stability)," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 648–654, Mar. 1999.

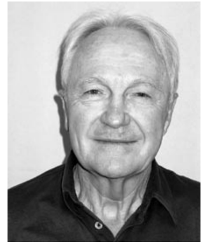
Christopher V. Rao received the B.S. degree from Carnegie Mellon University, Pittsburgh, PA, and the Ph.D. degree from the University of Wisconsin. Madison, both in chemical engineering, in 1994 and 2000, respectively.

He is currently a Research Associate with the Howard Hughes Medical Institute and a Postdoctoral Fellow at the University of California, Berkeley, and Lawrence Berkeley National Laboratory. His current research interests include genetic regulation, cell signaling, and microbial pathogenesis.



James B. Rawlings received the B.S. degree from the University of Texas, Austin, and the Ph.D. degree from the University of Wisconsin, Madison, both in chemical engineering, in 1979 and 1985, respectively.

He spent one year at the University of Stuttgart, Stuttgart, Germany, as a NATO Postdoctoral Fellow, and then joined the Faculty of the University of Texas. He moved to the University of Wisconsin, Madison, in 1995, where he is currently the Department Chair and Paul A. Elfers Professor in the Department of Chemical Engineering and the Codirector of the Texas-Wisconsin Modeling and Control Consortium (TWMCC). His research interests are in the areas of chemical process modeling and control, nonlinear model predictive control, moving horizon state estimation and monitoring, particle technology, and crystal engineering.



David Q. Mayne received the B.Sc. and M.Sc degrees in engineering from the University of the Witwatersrand, South Africa, the Ph.D. and D.Sc degrees from the University of London, U.K, and the degree of Doctor of Technology (*honoris causa*) from the University of Lund, Sweden, in 1950, 1955, 1967, 1981, and 1995, respectively.

He has held posts at the University of the Witwatersrand, Johannesburg, South Africa, the British Thomson Houston Company, Rugby, U.K., University of California, Davis, and Imperial College, London, U.K., where he is now a Senior Research Fellow. His research interests include optimization, optimization based design, nonlinear control, model predictive control, and adaptive control.