

Adaptive arrival cost update for improving Moving Horizon Estimation performance

G.Sánchez^{a,*}, M.Murillo^a, L.Giovanini^a

^aResearch Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL/CONICET, Ciudad Universitaria UNL, 4º piso FICH, (S3000) Santa Fe, Argentina.

Abstract

Moving horizon estimation is an efficient technique to estimate states and parameters of constrained dynamical systems. It relies on the solution of a finite horizon optimization problem to compute the estimates, providing a natural framework to handle bounds and constraints on estimates, noises and parameters. However, the approximation of the arrival cost and its updating mechanism are an active research topic. The arrival cost is very important because it provides a mean to incorporate information from previous measurements to the current estimates and it is difficult to estimate its true value. In this work, we exploit the features of adaptive estimation methods to update the parameters of the arrival cost. We show that, having a better approximation of the arrival cost, the size of the optimization problem can be significantly reduced guaranteeing the stability and convergence of the estimates. These properties are illustrated through simulation studies.

Keywords: Moving Horizon Estimation, Arrival cost, Adaptive estimation, Constrained state estimation, Variable forgetting.

1. Introduction

In control engineering, model-based control schemes assume that the states and parameters of the system are available for control law implementation. In practice, noisy measurements is the only information available from the system. Thus, the states and parameters have to be determined from these measurements using a dynamic model of the system. For linear systems, this problem has been solved and several methods based on different statistical measures have been developed [1, 2].

The physical limits of the system can be modeled through bounds on its states and parameters. The omission of such information in the estimation algorithm may substantially hamper its performance [3]. Unfortunately, Kalman filter can not handle explicitly bounds on estimates (states and parameters) and ad hoc methods have been developed to handle constraints [4]. The different approaches to enforce constraints in Kalman filter include model reduction [5], estimate projection [6], gain projection [7], probability density function truncation [8] and system projection [9]. These methods yield to suboptimal solutions at the best and unrealistic estimates when the statistics of unknown variables (initial states, measurement and process noises, disturbances) are poorly chosen. On the other hand, moving horizon estimation (MHE) solves an optimization problem to find the system estimates, providing a theoretical framework for constrained estimation.

MHE solves at each sample a finite horizon state estimation problem to determine the states and parameters of the system.

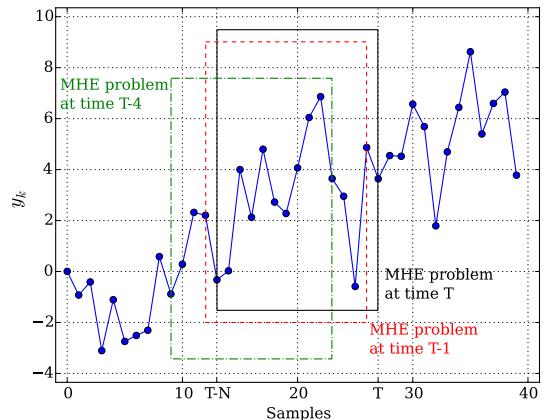


Figure 1: MHE smoothing update

When new measurements become available the old ones are discarded from the estimation window, and the estimation problem is solved to determine the new estimates (see Figure 1). The information of measurements which are not included in the estimation window is assimilated into the objective function through an extra term called *arrival cost*. It characterizes the statistical distribution of states at the beginning of the estimation window given the prior measurements information. In this way, it allows MHE to approximate the *full information* problem only considering a finite number of samples. A good approximation of the arrival cost allows to reduce the size of the estimation window and the size of the optimization problem, while it retains a good performance and robustness. The most accepted way of approximating the arrival cost is using a weighted 2-norm of the states at the beginning of the esti-

*Corresponding author

Email addresses: gsanchez@sinc.unl.edu.ar (G.Sánchez), mmurillo@sinc.unl.edu.ar (M.Murillo), lgiovanini@sinc.unl.edu.ar (L.Giovanini)

mation window [10, 11, 12, 13]. For linear systems Rao et al. [10] proposed to update the parameters of the arrival cost term (the weight matrix and initial states) using a Kalman filter or a Kalman smoother. The Gaussian distribution employed by these methods to model the conditional probabilities densities of initial states [10, 14] results in inadequate approximation of the arrival cost when estimates are constrained, leading to poor and unrealistic estimates. This problem arises from the fact that the presence of bounds on estimates modifies the probability distribution of noises and estimates, forcing to zero the probability of some values and eliminating the independence between estimates and noises [15].

To tackle these problems, Chu et al. [13] derived an iterative arrival cost update scheme that uses a quadratic approximation and information of active/inactive constraints from previous iteration. These ideas allow to build a quadratic approximation in the proximity of the optimal solution of the exact arrival cost. This update scheme is based on the hypothesis that the set of active constraints does not change after a particular time as the estimation horizon grows. This hypothesis works well when the estimation window is large, giving good results. However, if the hypothesis is not verified, i.e. some constraints become inactive after the state smoothing, the estimator may diverge. Furthermore, this updating mechanism overweights past data (by retaining active constraints) de-emphasizing the effect of information available in the new data on estimates. This fact can also cause divergence of the estimator if estimates are strongly correlated in time.

Finally, in a recent work Al-Matouq and Vincent [16] developed a MHE algorithm based on multiple estimation windows. The algorithm takes advantage of constraint inactivity to reduce the size of the optimization problem while it retains the stability and performance properties of full-information estimator. These properties are preserved by approximating the arrival cost with an unconstrained estimator that reformulates at each sample the objective function in the regions of constraints inactivity, allowing efficient long estimation windows.

In this work the weighting matrix is updated using adaptive estimation algorithms in combination with the MHE filter solution in the previous sample, while the initial states are the MHE filter solution in the previous sample. It should be noted that approximating the arrival in this way is consistent with constraints and bound on estimates and noises, while at the same time it introduces a feedback mechanism between measured data and estimates, improving the overall performance of the estimator. In this way, the weighting matrix is computed in a closed-loop fashion rather than in an open-loop way, as it is done in standard estimation techniques. The efficiency of the proposed approach is evaluated by conducting simulations studies on a benchmark problem available in the literature. The main contributions of this paper are to show that *i)* it is important to take into account bounds on estimates when the arrival cost is updated, *ii)* this idea can be implemented using adaptive estimation techniques to construct an approximation of the arrival cost, *iii)* the estimation resulting from a MHE with the proposed updating scheme is stable, and *iv)* the proposed updating scheme allows to shorten the estimation window without sacrificing per-

formance, reducing the size of the optimization problem. The paper is organized as follows: In Section 2 the MHE problem is presented. In Section 3 two different ways of updating the arrival cost are analysed. Simulation results are discussed in Section 5 and the conclusions are outlined in Section 6.

Notation: In the sequel $x \in \mathbb{R}^n$ is a column vector and its transpose is denoted as x^T , \mathbf{x}_k denotes a sequence of vectors over a given index up to k , for example $\mathbf{x}_k = \{x_j : j = 0, 1, \dots, k\}$, $A \in \mathbb{R}^{n \times m}$ is a $n \times m$ matrix and its transpose is denoted by A^T . If $A \in \mathbb{R}^{n \times n}$ its inverse is denoted by A^{-1} . Given a symmetric real matrix $P \in \mathbb{R}^{n \times n}$, it is said to be positive definite if for all $z \in \mathbb{R}^n$, $z^T P z > 0$. In the following we will use the short notation $P > 0$ to denote that matrix P is positive definite. By $\|\cdot\|$ we denote the Euclidean vector or induced matrix norm. For $x \in \mathbb{R}^n$ and $P > 0 \in \mathbb{R}^{n \times n}$, we let $\|x\|_P = \sqrt{x^T P x}$.

2. Problem formulation

Consider the linear time invariant discrete system

$$\begin{aligned} x_{k+1} &= Ax_k + Gw_k, \\ y_k &= Cx_k + v_k, \end{aligned} \quad (1)$$

where $x_k \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$ is the state vector, $w_k \in \mathcal{W} \subseteq \mathbb{R}^{n_w}$ is the state noise vector, $y_k \in \mathbb{R}^{n_y}$ is the measurement vector, $v_k \in \mathcal{V} \subseteq \mathbb{R}^{n_v}$ is the measurement noise vector and $k \in \mathbb{I}$. In the following, \hat{w}_k and \hat{v}_k are considered zero mean stationary stochastic disturbances with finite moments and the restriction sets \mathcal{X} , \mathcal{W} and \mathcal{V} are considered closed with $\mathbf{0} \in \mathcal{X}$, $\mathbf{0} \in \mathcal{W}$ and $\mathbf{0} \in \mathcal{V}$.

The full information estimator¹ (*FIE*) problem uses a sequence of measurements y_k to find the state estimates \hat{x}_k that solve the following optimization problem

$$\begin{aligned} \min_{\hat{x}_{jk}, \hat{w}_k} \Phi_k &= \|\hat{x}_{0|k} - \bar{x}_0\|_{P_0^{-1}}^2 + \sum_{j=0}^k \|\hat{w}_{j|k}\|_{Q^{-1}}^2 + \|\hat{v}_{j|k}\|_{R^{-1}}^2 \\ \text{s.t. } &\left\{ \begin{array}{l} \hat{x}_{j+1|k} = A\hat{x}_{jk} + \hat{w}_{jk}, \\ y_j = C\hat{x}_{jk} + \hat{v}_{jk}, \\ \hat{x}_{jk} \in \mathcal{X}, \hat{w}_{jk} \in \mathcal{W}, \hat{v}_{jk} \in \mathcal{V}, \end{array} \right. \end{aligned} \quad (2)$$

where Q^{-1} and R^{-1} are symmetric positive definite matrices that penalize the estimated noise vectors \hat{w}_{jk} and the output prediction \hat{v}_{jk} error, respectively, and the pair (\bar{x}_0, P_0^{-1}) summarizes the prior information at time $k = 0$ where \bar{x}_0 is the current knowledge of the initial estimate and P_0^{-1} is a symmetric positive definite weighting matrix. The solution of problem (2) yields smoothed estimates $\{\hat{x}_{jk} : j = 0, 1, \dots, k-1\}$, the filtered estimate $\hat{x}_{k|k}$ and sequence of estimated noise vectors, denoted as $\hat{\mathbf{w}}_k$ and $\hat{\mathbf{v}}_k$.

Since the full information estimator uses all measurements, as new ones become available the problem size grows with time making it intractable. MHE overcomes this problem by only considering a fixed amount of data and dynamic updates by sliding a window with time. The past measurements are taken

¹Other authors use the term batch least squares

into account through a penalty cost term Z_{k-N} called “arrival cost” [10]. Then, the MHE problem consists in finding the states, state and measurement noises that solve the following optimization problem

$$\begin{aligned} \min_{\hat{x}_{k-N|k}, \hat{w}_k} \Psi_k^N &= Z_{k-N}(\hat{x}_{k-N|k}) + \sum_{j=k-N}^k \|\hat{w}_{jk}\|_{Q^{-1}}^2 + \|\hat{v}_{jk}\|_{R^{-1}}^2 \\ \text{s.t. } &\left\{ \begin{array}{l} \hat{x}_{j+1|k} = A\hat{x}_{jk} + \hat{w}_{jk}, \\ y_j = C\hat{x}_{jk} + \hat{v}_{jk}, \\ \hat{x}_{jk} \in \mathcal{X}, \hat{w}_{jk} \in \mathcal{W}, \hat{v}_{jk} \in \mathcal{V}. \end{array} \right. \end{aligned} \quad (3)$$

The minimization of the full information problem (2) is equivalent to minimizing the fixed horizon problem (3) under the condition that Z_{k-N} is the exact arrival cost Z_{k-N}^*

$$\begin{aligned} Z_{k-N}^* &= \min_{\hat{x}_{0|k-N}, \hat{w}_{k-N}} \|\hat{x}_{0|k-N} - \bar{x}_0\|_{P_0^{-1}}^2 + \\ &\quad \sum_{j=0}^{k-N-1} \|\hat{w}_{jk-N}\|_{Q^{-1}}^2 + \|\hat{v}_{jk-N}\|_{R^{-1}}^2 \\ \text{s.t. } &\left\{ \begin{array}{l} \hat{x}_{j+1|k-N} = A\hat{x}_{jk-N} + \hat{w}_{jk-N}, \\ y_j = C\hat{x}_{jk-N} + \hat{v}_{jk-N}, \\ \hat{x}_{jk-N} \in \mathcal{X}, \hat{w}_{jk-N} \in \mathcal{W}, \hat{v}_{jk-N} \in \mathcal{V}. \end{array} \right. \end{aligned} \quad (4)$$

The exact arrival cost Z_{k-N}^* is a complicated piecewise quadratic function, whose transitions between segments are fixed upon the state of the constraints (active/inactive) in problem (2). Therefore, computing Z_{k-N}^* involves the solution of a constrained quadratic problem (QP) which size grows with the horizon $k - N$. Consequently, Z_{k-N}^* can not longer be computed solving the optimization problem (4) and an approximation \tilde{Z}_{k-N} is used. In conventional MHE formulations, a long estimation window is generally used to reduce the effect of the errors introduced by an improper approximation of the arrival cost [17]. In this way, MHE uses all the information available for the system, discarding the one that has no effect in the estimation due to the fading effect² (finite memory) of the system. On the other hand, if the estimation window is shortened, a better approximation of the arrival cost is needed such that it incorporates this information into the problem. For linear systems, it is generally assumed that the distribution is Gaussian and the weighting matrix in the arrival cost is updated using a Kalman filter [18]. Moreover, for unconstrained linear systems under the Gaussian assumption MHE is equivalent to the Kalman filter. The penalty matrix is generally calculated using dynamic programming techniques, and it turns out to be the covariance of the state estimate at time $k - N$ in the stochastic interpretation [10].

3. Arrival cost update

The MHE update relies on representing the full information problem (2) in the form (3), where N is fixed. Then, for each $k > N$, the arrival cost Z_{k-N}^* is computed using (4) and a new solution for the MHE problem is obtained. To avoid the curse

of dimensionality of problem (4), it can be updated recursively by extending the partial minimization of (4) to the next decision variable

$$\begin{aligned} Z_{k-N}^* &= \min_{\hat{w}_k} Z_{k-N-1}^* + \|\hat{w}_{k-N-1|k}\|_{Q^{-1}}^2 + \|\hat{v}_{k-N|k}\|_{R^{-1}}^2 \\ \text{s.t. } &\left\{ \begin{array}{l} \hat{x}_{k-N|k} = A\hat{x}_{k-N-1|k} + \hat{w}_{k-N-1|k}, \\ y_{k-N} = C\hat{x}_{k-N|k} + \hat{v}_{k-N|k}, \\ \hat{x}_{k-N|k} \in \mathcal{X}, \hat{w}_{k-N-1|k} \in \mathcal{W}, \hat{v}_{k-N|k} \in \mathcal{V}. \end{array} \right. \end{aligned} \quad (5)$$

Starting with

$$Z_0^* = \|\hat{x}_0 - \bar{x}_0\|_{P_0^{-1}}^2, \quad (6)$$

since it is hard to find an analytic expression for the exact arrival cost of the constrained MHE problem, we will follow Rao et al. [10] approach and use an approximate arrival cost based on the arrival cost for the unconstrained problem. Therefore, the exact arrival cost will be approximated by the following quadratic approximation

$$\tilde{Z}_{k-N} = \|\hat{x}_{k-N} - \bar{x}_{k-N}\|_{P_{k-N}^{-1}}^2, \quad (7)$$

where P_{k-N} and \bar{x}_{k-N} fully define the approximate arrival cost. The update of the approximate arrival cost \tilde{Z}_{k-N} can be formulated as an update of P_{k-N} and \bar{x}_{k-N} . Considering the partial minimization problem for the approximate iteration (5), where \tilde{Z}_{k-N-1} is given by (7), the result is a constrained sequential estimation problem.

For unconstrained problems, the solution of (5) using (7) leads to an unconstrained sequential estimation problem that can be solved using a Kalman filter [18]. Under this approach, the prior information is transferred to the current estimated window by conditioning the estimates of time k using the optimal estimate $\hat{x}_{k-N-1|k-N-1}^*$. For constrained problems this idea can still be employed, replacing the unconstrained Kalman filter by any of the ad-hoc methods described in [4]. However, the constraints are only applied to estimates and the weighting matrices remain unaffected. The major argument in favour of this kind of update lays in the fact that no input information y_j is over weighted during the estimation process, which guarantees that no measurement information is used twice. One disadvantage of using the filtering update lays in the fact that a periodic behavior associated with relying on estimates based on few data points occurs. This phenomenon is known as *estimation cycling* and can be avoided using the smoothing update [19]. The smoothing formulation takes advantage of more information by including more data in the \bar{x}_{k-N} update. Another problem is the inaccurate approximation of the arrival cost when inequality constraints switch between being active and inactive at the optimal solution. Constraints modify the probability distribution and noises in a way that analytical updates can not properly approximate. In this work, the approximate arrival cost update (5) is formulated as an update of P_{k-N} and \bar{x}_{k-N} in (7) in two steps: *i*) update \bar{x}_{k-N} and then *ii*) update P_{k-N} .

3.1. Computing \bar{x}_{k-N}

One of the components of \tilde{Z}_{k-N} is the initial state \bar{x}_{k-N} . Since the MHE problem yields smoothed estimates $\{\hat{x}_{jk} : j = k -$

²With fading effect we mean that if the system can be represented with a FIR model of order N , then an horizon of size N is employed, thus the system is fully described inside the horizon.

$N, \dots, k-1\}$ and the filtered estimate $\hat{x}_{k|k}$, we can use the one time-step before optimal smoothed state estimate $\hat{x}_{k-N|k-1}^*$,

$$\bar{x}_{k-N} = \hat{x}_{k-N|k-1}^*.$$

The use of the smoothed update implies the use of similar approximation hypothesis like Chu et al. [13]: the state of constraints (active/inactive) does not change once the estimate leaves the estimation window.

3.2. Computing P_{k-N}

The weighting matrix P_{k-N} can be seen as the covariance of x_{k-N} in the stochastic interpretation of the MHE problem [10]. Analytical solutions for computing P_{k-N} only exists for linear unconstrained problems. For linear constrained problems (and nonlinear problems) approximate solutions have been derived. They are based on a recursive update that employs the information of the system available at time k (system model, previous covariance matrix, active constraints, covariance of the noises). These equations evolve in an open-loop fashion, based on the system model and the assumed statistics of the noises, without incorporating any information from measurements or estimates.

In signal processing and adaptive estimation a recursive update based on measurements and/or estimates is often used [20]

$$P_{k-N}^{-1} = \alpha_k P_{k-N-1}^{-1} + \beta_k \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T \quad P_0 > 0, \quad (8)$$

where $0 < \alpha_k < 1$ and $0 < \beta_k \leq 2 \forall k \geq 0$ are time-varying weighting sequences. Equation (8) generates recursively a real-time estimation of the weighting matrix P_{k-N} by updating the previous estimates P_{k-N-1} with an exponential time averaging of $\hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T$. This updating mechanism can be understood as a time-varying filter whose input is $\hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T$ and the initial condition is P_0 .

Lemma 3.1. Let $\{A_i : i = 1, \dots, n\}$ be positive definite and let $\{\gamma_i : i = 1, \dots, n\}$ be nonnegative real numbers. Then $\sum_{i=1}^n \gamma_i A_i$ is positive semidefinite.

Proof. Let $x \in \mathbb{R}^n$ be nonzero and observe that $x^T (\sum_{i=1}^n \gamma_i A_i) x = \sum_{i=1}^n \gamma_i (x^T A_i x) \geq 0$. Since each $\gamma_i \geq 0$ and each $x^T A_i x > 0$. The latter sum results in a positive definite matrix if any of the summands is positive. \square

Lemma 3.2. If $P_0 > 0$ and $\alpha_0 > 0$, then the matrix P_k obtained using Eq. (8) is positive definite for all $k \geq 0$.

Proof. Since xx^T is positive semidefinite and from Lemma 3.1 it follows that P_k is positive definite for all $k \geq 0$. \square

Lemma 3.3. If $P > 0$, then there exists an invertible matrix Z such that $Z^T P Z = I$ and $Z Z^T = P^{-1}$.

Proof. See Lemma 7, page 261 from Johnson [21]. \square

Note that α_k and β_k in (8) have opposite effects. When $\alpha_k \ll 1$ the adaptation algorithm tends to discard the old information summarized in P_{k-N-1} , estimating P_{k-N} using the recent data $\hat{x}_{k-N|k-1}$. This fact allows to avoid the overweighting of past data (by retaining data with active constraints) that

can deteriorate the estimator performance, improving the estimation of $\hat{x}_{k-N|k}$. From an optimization point of view, when the algorithm discards the data without useful information, the matrix P_{k-N}^{-1} increases and the initial information available in the arrival cost tends to be more relevant. On the contrary, when $\alpha_k \approx 1$ the adaptation algorithm tends to retain the old data summarized in P_{k-N-1} . This fact allows to include all the past information that is relevant for the estimation problem, improving the estimation of $\hat{x}_{k-N|k}$ by including more information and forgetting only initial data, facts that decrease P_{k-N}^{-1} . When $\beta_k \gg 0$ the adaptation algorithm tends to increase P_{k-N}^{-1} by incorporating information from estimates $\hat{x}_{k-N|k-1}$. This fact allows to include more information from recent data into P_{k-N}^{-1} , increasing its value. When $\beta_k \approx 0$ the adaptation algorithm discards the information available in $\hat{x}_{k-N|k-1}$ and only applies a forgetting mechanism on P_{k-N}^{-1} , which depends on the value of α_k . In general, α_k and β_k allow the estimation algorithm to discard and/or incorporate information into P_{k-N}^{-1} according to the information available in the measurements $\{y_j\}_{j=k-N}^k$, estimates $\hat{x}_{k-N|k}$ and process noises $\{w_j\}_{j=k-N}^k$.

Since the existence of the inverse is guaranteed by Lemma 3.3, we can use the matrix inversion lemma to rewrite Eq. (8) as follows

$$P_{k-N} = \frac{1}{\alpha_k} \left[P_{k-N-1} - \frac{P_{k-N-1} \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T P_{k-N-1}}{\frac{\alpha_k}{\beta_k} + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}} \right]. \quad (9)$$

This way of computing P_{k-N} introduces a feedback loop into the MHE problem. This fact allows the MHE problem to adjust P_{k-N}^{-1} according to the information available in the measurements $\{y_j\}_{j=k-N}^k$ and to guarantee the convergence of estimates $\hat{x}_{k-N|k}$ and $\{\hat{w}_j\}_{j=k-N}^k$. We must be careful in the choice of the arrival cost update mechanism, since poor approximations can lead to instability [22, 23] in the estimator. In the following Sections the conditions that sequences α_k and β_k must satisfy to guarantee convergence will be studied. If the current sequence of process noise $\{w_j\}_{j=k-N}^k$ is large, P_{k-N}^{-1} will grow emphasizing \tilde{Z}_{k-N} over the other terms. This fact will lead to $\tilde{Z}_{k-N} \rightarrow 0$ and improving the overall estimation. Then, P_{k-N}^{-1} will decrease until \tilde{Z}_{k-N} achieves its stationary value or the feedback starts again (see Fig 3b).

Depending on the values of α_k and β_k , different forgetting factor profiles, suitable for different adaptation problems, are obtained [20]. For example, if the estimates are varying slowly and constraints does not change their state after leaving the estimation window, the most appropriate selection is

$$0 < \alpha_k = \alpha < 1, \quad \beta_k = 1, \quad (10)$$

which leads to a *constant forgetting factor*. In this case, equation (9) will estimate P_{k-N} using the available MHE estimates, de-emphasizing past observations at a constant rate. When a constant forgetting factor is used, the choice of α_k and β_k given by Eq (10) leads to a decreasing adaptation gain, where old information is continually forgotten. If the forgetting factor α is not carefully chosen in this schema, this may lead to an exponential growth of the matrix P_{k-N} and it can turn the MHE

algorithm extremely sensitive to disturbances and susceptible to numerical difficulties [24].

This is why several researchers [25, 26, 27] suggest to use a variable forgetting factor, which enables the algorithm to follow slow and sudden changes in the system, while it can prevent matrix P_{k-N} from blowing-up. This choice is appropriate when the estimates are stationary and constraints can change their state after leaving the estimation window [20], since some forgetting must be done in order to maintain the information of the relevant constrained states. In this work we review some of the many mechanisms available to update the parameters α_k and β_k . For example, if the evolution of α_k is given by

$$\begin{aligned} \alpha_k &= 1 - \frac{\hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}}{1 + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}}, \\ \beta_k &= 1. \end{aligned} \quad (11)$$

We can see that the forgetting factor in Eq (11) depends on estimates, leading to a *variable forgetting factor*. It automatically takes the value 1 if the norm of the states becomes null, leading to the standard Recursive Least Squares (RLS) algorithm. In the cases where the states sequence is such that the term $\hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}$ is significant with respect to one, the forgetting factor takes a lower value assuring good adaptation capabilities. The algorithm will provide a quick adaptation and will prevent the exponential growth of P_{k-N} , stabilizing the estimates (controlling the amount of information employed to compute P_{k-N}) and the MHE problem (3).

If the estimates are time-varying and constraints can change after leaving the estimation window, the forgetting factor has to guarantee a constant trace of the weighting matrix, which can be obtained by

$$\begin{aligned} \alpha_k &= \frac{1}{\Xi} \text{trace} \left(P_{k-N-1} - \frac{\hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}}{\eta + \hat{x}_{k-N|k-1}^T P_{k-N} \hat{x}_{k-N|k-1}} \right), \\ \beta_k &= \eta^{-1} \alpha_k, \end{aligned} \quad (12)$$

where $\Xi > 0$ is the desired trace of the weighting matrix and $\eta > 0$ is constant.

Fortescue et al. [24] proposed a variable forgetting algorithm that tries to keep the amount of information constant and shows that a reasonable choice of information measure can prevent the weighting matrix from blowing-up, while retaining the adaptability. Later, Osorio Cordero and Mayne [26] modified that algorithm in order to ensure global convergence. The modifications ensure that $\text{trace } P_k$ (and, hence $\|P_k\|$) is always less than a given constant c . The algorithm yields the standard least-squares estimate if $\alpha_k = 1$, and the weighted least-squares estimate if $\alpha_k \in (0, 1)$. The memory length N_k is related to the forgetting factor α_k through $\alpha_k = 1 - 1/N_k$; forgetting factors of 0, 0.99 and 1 correspond to memory lengths of 1, 100 and ∞ , respectively. The various constant required by the algorithm may be chosen as follows: σ is chosen so that σ/σ_w is large, where σ_w denotes the process noise variance; $P_0 = \delta I$, where δ is large (10^6 , for example), $c > \delta$ is large. The variable forgetting factor and the information matrix can be calculated as

follows

$$\begin{aligned} \epsilon_k &= y_{k-N} - \hat{y}_{k-N} \\ N_k &= [1 + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}] \frac{\sigma}{\|\epsilon_{k-N}\|_2^2} \\ \alpha_k &= 1 - 1/N_k \\ W_k &= [I - \frac{P_{k-N-1} \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T}{1 + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}}] P_{k-N-1} \\ P_{k-N} &= \frac{1}{\alpha_k} W_k, \text{ if } 1/\alpha_k \text{ trace}(W_k) \leq c \\ P_{k-N} &= W_k, \text{ otherwise.} \end{aligned} \quad (13)$$

Remark 1. The weighting matrix estimation algorithm can be recasted like a Kalman filter with the smoothed estimate $\hat{x}_{k-N|k-1}$ as estimated states, \hat{v}_k as the estimation residual and P_{k-N} the covariance matrix [20]. In case of unconstrained estimation, the weighting matrix estimation algorithm is equivalent to a filter update [18]. The relationship between RLS and the Kalman filter is well known [1], where in the case of the Kalman filter, the weighting matrix update formula is a special case of Eq. (8) under specific assumptions and a certain choice of α_k and β_k . For a detailed description the reader can see Section 3.2.4 of Landau et al. [20].

3.3. MHE update summary

The MHE approach solves problem (3), providing a solution $\{\hat{x}_{j|k}\}_{j=k-N}^k$ that approximates the optimal solution of problem (2). For $k \leq N$, the full information problem (2) is solved. Then, for $k > N$ the MHE problem (4) is solved with the approximate arrival cost

$$\tilde{Z}_{k-N} = \|\hat{x}_{k-N|k} - \hat{x}_{k-N|k-1}^*\|_{P_{k-N}^{-1}}^2, \quad (14)$$

whose solution can be obtained using any suitable QP solver. Then, P_{k-N} and \bar{x}_{k-N} are updated and the horizon is receded, updating the measurements vector $\{y_j\}_{j=k-N}^k$. Algorithm 1 summarizes the proposed moving horizon estimator. The method used for the adaptive weighting matrix update defines the algorithm behaviour. If we wish to use a constant trace approach (MHE_{AD-CT}), Eq. (12) is used. On the other hand, if we wish to use a variable forgetting factor (MHE_{AD-VF}), Eq. (13) is used.

4. Stability analysis

Stability of the estimator implies that the reconstruction error converges to zero for the nominal system

$$\begin{aligned} x_{k+1} &= Ax_k, \\ y_k &= Cx_k, \end{aligned} \quad (15)$$

with no state or measurement noise. However, for the constrained case, a poor choice of constraints may prevent convergence to the true state of the system. Rao et al. [10] added the requirement that the evolution of the system respects the constraints, which need to satisfy only the following weaker assumption to prove stability.

Algorithm 1: MHE_{AD} Algorithm

$\bar{x}_0, P_0 > 0, Q, R > 0$ and $N \geq 1$

Initialization:

$\bar{x}_{k-N} \leftarrow \bar{x}_0$

$P_{k-N} \leftarrow P_0$

for samples $k = 0, 1, 2, \dots$ **do**

 Obtain measurement y_k

if $k < N$ **then**

 Prepare data vector:

$$Y \leftarrow [y_0, \dots, y_k]^T$$

 Solve Φ_k with \bar{x}_0, P_0 and Y

else

 Prepare data vector:

$$Y \leftarrow [y_{k-N}, \dots, y_k]^T$$

 Solve Ψ_k^N with \bar{x}_{k-N}, P_{k-N} and Y

$$\bar{x}_{k-N} \leftarrow \hat{x}_{k-N+1|k}$$

 Update P_{k-N} using Eq. (12) or (13)

end if

 Obtain current estimates $\hat{x}_{j|k}$ for $j = k - N, \dots, k$

end for

Assumption 1. Suppose the system (15) with initial condition x_0 generates the data. We assume there exists $x_{0|\infty}, \{w_{k|\infty}\}_{k=0}^\infty$ and $\sigma > 0$ such that

$$\sum_{k=0}^{\infty} \|v_{k|\infty}\|_{R^{-1}} + \|w_{k|\infty}\|_{Q^{-1}} + \|x_{0|\infty} - \hat{x}_0\|_{P_0^{-1}} \leq \sigma \|x_0 - \hat{x}_0\|^2$$

and $x_{k|\infty} \in \mathcal{X}, w_{k|\infty} \in \mathcal{W}, v_{k|\infty} \in \mathcal{V}$.

Assumption 1 states the existence of a feasible state and disturbance trajectory that yields bounded cost. Next we state the stability results for the FIE.

Remark 2. Suppose the matrices Q , R , and P_0 are positive definite, (C, A) is observable, and assumption 1 holds. Then, the constrained full information estimator is an asymptotically stable observer for the system given by Eq. (15).

Proof. See Muske et al. [18]. \square

Remark 3. The matrix P_k satisfies $\|P_k\| \leq c, \forall k$. The boundedness of P_k guarantees that the optimization problem for MHE with adaptive arrival cost update is convex and that under Assumption 1, a solution exists.

Proof. From Lemma 3.2, for all k , P_k is symmetric and positive definite. Its norm $\|P_k\|$ is equal to $\lambda_{\max}(P_k)$, the maximum eigenvalue of P_k . Its trace satisfies $\|P_k\| \leq \text{trace}(P_k)$.

By choice $\|P_0\| \leq \text{trace}(P_0) < c$. Suppose $\|P_k\| \leq c$. From Eq. (13), if $(1/\alpha_k)\text{trace}(W_k) \leq c$, then $\|P_k\| = (1/\alpha_k)\|W_k\| \leq (1/\alpha_k)\text{trace}(W_k) \leq c$. If, on the other hand, $(1/\alpha_k)\text{trace}(W_k) > c$, then $\alpha_k = 1$ and $P_k^{-1} = P_{k-1}^{-1} + \hat{x}_{k-N|k-1}\hat{x}_{k-N|k-1}^T$, i.e. P_k is less positive definite than P_{k-1} , so that $\|P_k\| \leq \|P_{k-1}\| \leq c$. Hence, $\|P_k\| \leq c$. \square

Remark 4. The matrix P_k has a steady state-solution,

$$\lim_{k \rightarrow \infty} P_k = P_\infty$$

Proof. The update formula for P_{k-N} given by Eq. (13) can be seen as a special case of a Riccati equation as follows

$$\begin{aligned} \alpha_k P_{k-N} &= P_{k-N-1} - \frac{P_{k-N-1} \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T P_{k-N-1}}{1 + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}} \\ &= P_{k-N-1} - \\ &\quad P_{k-N-1} \hat{x}_{k-N|k-1} (1 + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1})^{-1} \hat{x}_{k-N|k-1}^T P_{k-N-1} \end{aligned}$$

which is stable if and only if all of the eigenvalues of the closed loop state transfer matrix

$$I - \frac{\hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T P_{k-N-1}}{1 + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}} \quad (16)$$

are strictly inside the unit circle of the complex plane. The eigenvalues of the right hand side of the previous expression lie inside the unit circle, since

$$\|\hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T P_{k-N-1}\| \leq \|\hat{x}_{k-N|k-1}\| \|\hat{x}_{k-N|k-1}^T\| \|P_{k-N-1}\|$$

and

$$\|\hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}\| \leq \|\hat{x}_{k-N|k-1}^T\| \|P_{k-N-1}\| \|\hat{x}_{k-N|k-1}\|$$

hence

$$\left\| \frac{\hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T P_{k-N-1}}{1 + \hat{x}_{k-N|k-1}^T P_{k-N-1} \hat{x}_{k-N|k-1}} \right\| \leq 1. \quad (17)$$

Since the product of symmetric positive semidefinite matrixes is another symmetric positive semidefinite matrix, the eigenvalues of (17) are all greater than 0 and bounded by 1. Then, the eigenvalues of (16) satisfy that are greater than 0 and bounded by 1. Hence, the matrix P_k has a steady-state solution P_∞ . \square

Theorem 4.1. $V_k = \tilde{x}_k^T P_k^{-1} \tilde{x}_k$ is a Lyapunov function for MHE with adaptive arrival cost update, where $\tilde{x}_k = x_k - \hat{x}_k$ and the matrix P_k^{-1} is given by Eq. (13), i.e. $P_k^{-1} = \alpha_k(P_{k-1}^{-1} + \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T)$.

Proof. Since $\alpha_k V_{k-1} \leq V_{k-1}$, we can subtract $\alpha_k V_{k-1}$ to V_k

$$\begin{aligned} V_k - \alpha_k V_{k-1} &= \tilde{x}_k^T P_k^{-1} \tilde{x}_k - \alpha_k \tilde{x}_{k-1}^T P_{k-1}^{-1} \tilde{x}_{k-1} \\ &= (A \tilde{x}_{k-1})^T P_k^{-1} (A \tilde{x}_{k-1}) - \alpha_k \tilde{x}_{k-1}^T P_{k-1}^{-1} \tilde{x}_{k-1} \\ &= \tilde{x}_{k-1}^T A^T P_k^{-1} A \tilde{x}_{k-1} - \alpha_k \tilde{x}_{k-1}^T P_{k-1}^{-1} \tilde{x}_{k-1} \\ &= \tilde{x}_{k-1}^T A^T [\alpha_k (P_{k-1}^{-1} + \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T)] A \tilde{x}_{k-1} - \\ &\quad \alpha_k \tilde{x}_{k-1}^T P_{k-1}^{-1} \tilde{x}_{k-1} \\ &= \alpha_k \tilde{x}_{k-1}^T [A^T P_{k-1}^{-1} A + \\ &\quad A^T \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T A - P_{k-1}^{-1}] \tilde{x}_{k-1} \end{aligned}$$

Since P_k is symmetric positive definite for all k and $A^T \hat{x}_{k-N|k-1} \hat{x}_{k-N|k-1}^T A$ is symmetric positive semidefinite for all k , it follows that if the system given by Eq. (15) is observable, then MHE with adaptive arrival cost update is an asymptotically stable observer. \square

4.1. Bounded stability

From Remark 3 we know that $\|P_k\|$ is nonincreasing, hence $\|P_k^{-1}\|$ is nondecreasing. By the convergence of P_k to P_∞ , we get that $\|P_k^{-1}\| \leq \|P_{k+1}^{-1}\| \leq \|P_\infty^{-1}\|$. Therefore, the arrival and stage cost satisfy the following bounds

$$\lambda_{\min}(P_{k-N}^{-1})\|\hat{x}_{k-N|k}-\bar{x}_{k-N}\| \leq \|\hat{x}_{k-N|k}-\bar{x}_{k-N}\|_{P_{k-N}^{-1}} \leq \lambda_{\max}(P_\infty^{-1})\|\hat{x}_{k-N|k}-\bar{x}_{k-N}\|$$

$$\lambda_{\min}(Q^{-1})\|\hat{w}\| \leq \|\hat{w}\|_{Q^{-1}} \leq \lambda_{\max}(Q^{-1})\|\hat{w}\|$$

$$\lambda_{\min}(R^{-1})\|\hat{v}\| \leq \|\hat{v}\|_{R^{-1}} \leq \lambda_{\max}(R^{-1})\|\hat{v}\|.$$

Under bounded disturbances, the MHE estimator with adaptive arrival cost update is robust globally asymptotically stable. Furthermore, if $\lim_{k \rightarrow \infty} w_k = 0$ and $\lim_{k \rightarrow \infty} v_k = 0$, then $\lim_{k \rightarrow \infty} \|\hat{x}_k - \bar{x}_k\| = 0$.

Proof. See Theorem 7 of [28]. \square

5. Results

For the simulations tests we adopted the discrete-time model used by Rao et al. in [29]

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 0.99 & 0.2 \\ -0.1 & 0.3 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_k, \\ y_k &= \begin{bmatrix} 1 & -3 \end{bmatrix} x_k + v_k, \end{aligned} \quad (18)$$

where $\{v_k\}$ is a sequence of independent, normally distributed random variables with zero mean and covariance σ_v , and $w_k = |z_k|$ where $\{z_k\}$ is a sequence of independent, normally distributed random variables with zero mean and covariance σ_w . For all the following examples, the initial state \bar{x}_0 is normally distributed with zero mean and covariance $\sigma_x = 0.5$ and we assume that $x_0 = [0, 0]^T$ and $\bar{x}_0 = [0.5, -0.5]^T$. The constrained estimation problem is formulated here with $Q = 1$, $R = 0.01$, $P_0 = 0.5$ and we add the inequality constraint $w_k \geq 0$ to include the available information on the random sequence.

First, the proposed algorithms (MHE_{AD-VF} and MHE_{AD-CT}) are compared with the full information estimator (FIE) and with a linear algorithm (MHE_{KF}) available in the Nonlinear Model Predictive Control Tools for CasADI (mpc-tools-casadi) toolbox [30]. This algorithm propagates the arrival cost and initial state estimate using a Kalman filter. The simulations evaluates the effect of the horizon size on the performance of the estimators. Three different horizon sizes ($N = 3$, $N = 6$ and $N = 10$) were used for each MHE algorithm with $\sigma_v = 0.1$ and $\sigma_w = 1.0$. The sum square estimation error

$$\xi^{(i)} = \sum_{j=0}^k (x_j^{(i)} - \hat{x}_j^{(i)})^2,$$

where the superscript (i) denotes the i th component of x , was used as benchmark. In this test, 100 trials were run and the average sum square estimation error was computed. The results are shown in Tables 1 and 2. It can be seen that MHE_{AD} errors are always lower than MHE_{KF} and closer to FIE on every

	x ⁽¹⁾			
	FIE	MHE _{AD-VF}	MHE _{AD-CT}	MHE _{KF}
$N = 3$	13.22	20.44	27.37	103.42
$N = 6$	13.22	17.77	16.25	50.68
$N = 10$	13.22	15.28	14.51	26.54

Table 1: Comparison of the effect of the horizon size on the average sum square estimation error for state $x^{(1)}$.

	x ⁽²⁾			
	FIE	MHE _{AD-VF}	MHE _{AD-CT}	MHE _{KF}
$N = 3$	1.55	2.20	2.84	11.55
$N = 6$	1.55	1.96	1.76	5.71
$N = 10$	1.55	1.71	1.61	3.02

Table 2: Comparison of the effect of the horizon size on the average sum square estimation error for state $x^{(2)}$.

window size considered. As the window size is enlarged, the arrival cost plays a less significant role on the cost function of the optimization problem and the cost of the three estimators tends to converge to the same value. A good approximation of the arrival cost allows us to decrease the window size. In this aspect, it can be seen that MHE_{AD} errors do not change significantly with the different horizon sizes, showing that this algorithm seems to be a good method for the arrival cost approximation, even for short horizons.

5.1. Example 1: $\sigma_w = 1.0$, $\sigma_v = 0.1$, $N = 5$

In this example, the process noise covariance is set to $\sigma_w = 1.0$ and the measurement noise covariance to $\sigma_v = 0.1$. Figure 2 shows the states $x^{(1)}$, $x^{(2)}$ and its estimates (see Fig. 2a and 2b) with $N = 5$. It can be seen that MHE_{AD} estimation is closer to FIE despite the short horizon chosen. As time passes, the state estimates of MHE_{AD} converge to the estimates provided by FIE. Figures 2c and 2d show the square error $\xi^{(i)}$ at every sample time. In these figures it can be seen that the error of the proposed methods are similar to the one obtained using FIE. On the other hand, from these figures it can also be seen that the performance of MHE_{KF} is degraded, mainly because the window size was set to $N = 5$. This difference can be explained by the fact that KF update does not approximate properly the arrival cost due to the effect of constraints [15]. This situation gets worse when the horizon is shorter, as was mentioned in the previous experiment.

5.2. Example 2: $\sigma_w = 0.4$, $\sigma_v = 0.1$, $N = 5$

In this example, although the process noise covariance is set to $\sigma_w = 0.4$, it varies with time and the different MHE algorithms do not have knowledge of this happening. The covariance σ_w will take values 0.1 when $0 \leq k < 20$, 0.25 when $20 \leq k < 40$, 1.5 when $40 \leq k < 100$, 1.0 when $100 \leq k < 150$ and 0.4 when $150 \leq k < 200$. We can see in Fig. 3a that the

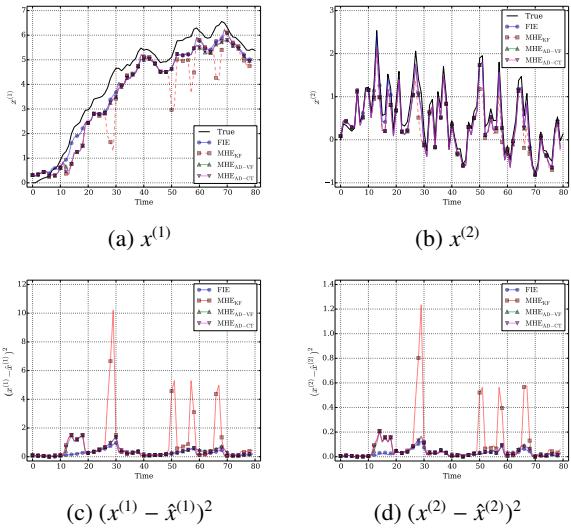


Figure 2: Comparison of estimators for model (18) with horizon size $N = 5$.

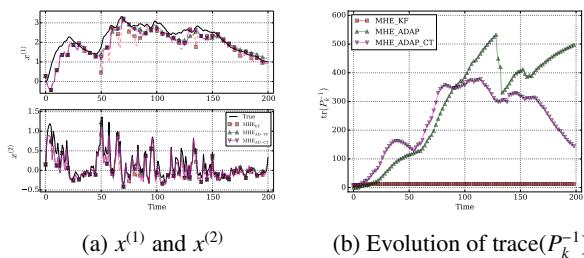


Figure 3: Comparison of estimators for model (18) with horizon size $N = 5$ and variable σ_w .

proposed algorithms still perform better than MHE_{KF}. Fig. 3b shows the trace of the matrix P_{k-N}^{-1} used by every algorithm except for FIE, since in that case the matrix P_0 does not evolve with time. We can see that the MHE_{KF} matrix converges to a steady state value and never changes again, while the proposed algorithms adapt its weight according to the changing conditions of the process noise. It can be seen that the trace of P_{k-N}^{-1} grows when the error is smaller, showing that it has more confidence on the estimates. On the other hand, when the difference between the real state and the estimate grows, the trace of the weighting matrix becomes smaller, providing better adaptation even under the influence of unknown and varying disturbances. In Figures 4a and 4b we can see the evolution of the forgetting factors of MHE_{VF} and MHE_{CT}, respectively. In both cases we can see that α_k is nearly 1 and the algorithm behaves like a weighted least-squares estimator with very long asymptotic memory length. In the cases where the noise increases and the estimator is not following the system, the forgetting factor decreases in order to allow for rapid adaptation of the weighting matrix. This is the main reason why the proposed algorithm provides a better approximation of the arrival cost.

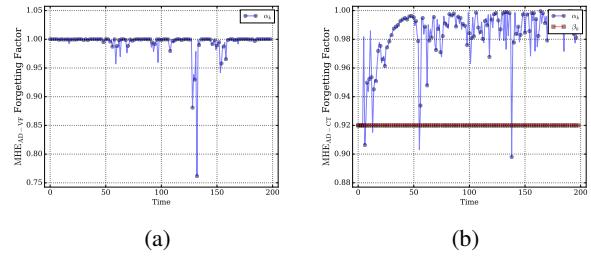


Figure 4: Evolution of the sequences of forgetting factors for MHE_{VF} and MHE_{CT}.

6. Conclusion

In this work we investigated an alternative method to approximate the arrival cost for the MHE problem on linear systems. To the best of our knowledge, adaptive estimation techniques have not been used before to update the weighting matrix P_{k-N} . From the results obtained we could show that adaptive estimation techniques give a good approximation of the arrival cost. This fact is due to the ability of the variable forgetting factor to provide good adaptation capabilities allowing the estimator to incorporate the relevant information of the data and follow not only slow changes but also sudden changes in the dynamics. It is worth mentioning that the proposed approach can be easily extended to nonlinear systems, since the update mechanism only depends on the state estimates and not on the model. The use of CasADi toolbox [31] allows for fast implementation of the optimization algorithms, leading to real-time operation.

7. Acknowledgments

The authors wish to thank Dr. Leandro Di Persia for the insightful comments on covariance estimation. The research was supported by *Universidad Nacional de Litoral* (with CAID 501201101 00529 LI) and *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET) from Argentina.

References

- [1] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Vol. 64, Academic Press, 1970.
- [2] J. L. Crassidis, J. L. Junkins, Optimal estimation of dynamic systems, CRC press, 2011.
- [3] E. L. Haseltine, J. B. Rawlings, Critical evaluation of extended kalman filtering and moving-horizon estimation, Industrial & engineering chemistry research 44 (8) (2005) 2451–2460.
- [4] D. Simon, Kalman filtering with state constraints: a survey of linear and nonlinear algorithms, Control Theory & Applications, IET 4 (8) (2010) 1303–1318.
- [5] D. Simon, T. L. Chia, Kalman filtering with state equality constraints, Aerospace and Electronic Systems, IEEE Transactions on 38 (1) (2002) 128–136.
- [6] D. L. Hall, S. A. McMullen, Mathematical techniques in multisensor data fusion, Artech House, 2004.
- [7] B. O. S. Teixeira, J. Chandrasekar, H. J. Palanthandlam-Madapusi, L. A. B. Torres, L. A. Aguirre, D. S. Bernstein, Gain-constrained kalman filtering for linear and nonlinear systems, Signal Processing, IEEE Transactions on 56 (9) (2008) 4113–4123.

- [8] D. Simon, D. L. Simon, Constrained kalman filtering via density function truncation for turbofan engine health estimation, International Journal of Systems Science 41 (2) (2010) 159–171.
- [9] S. Ko, R. R. Bitmead, State estimation for linear systems with state equality constraints, Automatica 43 (8) (2007) 1363–1368.
- [10] C. V. Rao, J. B. Rawlings, J. H. Lee, Constrained linear state estimation a moving horizon approach, Automatica 37 (10) (2001) 1619–1628.
- [11] C. C. Qu, J. Hahn, Computation of arrival cost for moving horizon estimation via unscented kalman filtering, Journal of Process Control 19 (2) (2009) 358–363.
- [12] P. Kühl, M. Diehl, T. Kraus, J. P. Schlöder, H. G. Bock, A real-time algorithm for moving horizon state and parameter estimation, Computers & chemical engineering 35 (1) (2011) 71–83.
- [13] E. Chu, A. Keshavarz, D. Gorinevsky, S. Boyd, Moving horizon estimation for staged qp problems, in: Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, IEEE, 2012, pp. 3177–3182.
- [14] C. V. Rao, J. B. Rawlings, Constrained process monitoring: Moving-horizon approach, AIChE journal 48 (1) (2002) 97–109.
- [15] D. G. Robertson, J. H. Lee, On the use of constraints in least squares estimation and control, Automatica 38 (7) (2002) 1113–1123.
- [16] A. A. Al-Matouq, T. L. Vincent, Multiple window moving horizon estimation, Automatica 53 (2015) 264–274.
- [17] S. C. Patwardhan, S. Narasimhan, P. Jagadeesan, B. Gopaluni, S. L. Shah, Nonlinear bayesian state estimation: A review of recent developments, Control Engineering Practice 20 (10) (2012) 933–953.
- [18] K. R. Muske, J. B. Rawlings, J. H. Lee, Receding horizon recursive state estimation, in: American control conference, Vol. 30, 1993, pp. 900–904.
- [19] M. J. Tenny, J. B. Rawlings, Efficient moving horizon estimation and nonlinear model predictive control, in: American Control Conference, 2002. Proceedings of the 2002, Vol. 6, IEEE, 2002, pp. 4475–4480.
- [20] I. D. Landau, R. Lozano, M. M'Saad, A. Karimi, Adaptive control, Vol. 51, Springer Berlin, 1998.
- [21] C. R. Johnson, Positive Definite Matrices, The American Mathematical Monthly 77 (3) (1970) 259. doi:10.2307/2317709.
 URL <http://www.jstor.org/stable/2317709?origin=crossref>
- [22] P. K. Findiesen, Moving horizon state estimation of discrete time systems, Master's thesis, University of Wisconsin, Madison (1997).
- [23] C. V. Rao, Moving horizon strategies for the constrained monitoring and control of nonlinear discrete-time systems, Ph.D. thesis, University of Wisconsin, Madison (2000).
- [24] T. Fortescue, L. Kershenbaum, B. Ydstie, Implementation of self-tuning regulators with variable forgetting factors, Automatica 17 (6) (1981) 831–835.
- [25] K. J. Åström, B. Wittenmark, On self tuning regulators, Automatica 9 (2) (1973) 185–199.
- [26] A. Osorio Cordero, D. Mayne, Deterministic convergence of a self-tuning regulator with variable forgetting factor, IEE Proceedings D Control Theory and Applications 128 (1) (1981) 19–23. doi:10.1049/ip-d.1981.0004.
 URL <http://digital-library.theiet.org/content/journals/10.1049/ip-d.1981.0004>
- [27] L. Ljung, S. Gunnarsson, Adaptation and tracking in system identificationA survey, Automatica 26 (1) (1990) 7–21. doi:10.1016/0005-1098(90)90154-A.
 URL <http://linkinghub.elsevier.com/retrieve/pii/000510989090154A>
- [28] M. A. Muller, Nonlinear moving horizon estimation for systems with bounded disturbances, in: 2016 American Control Conference (ACC), IEEE, 2016, pp. 883–888. doi:10.1109/ACC.2016.7525026.
 URL <http://ieeexplore.ieee.org/document/7525026/>
- [29] C. V. Rao, J. B. Rawlings, D. Q. Mayne, Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations, Automatic Control, IEEE Transactions on 48 (2) (2003) 246–258.
- [30] M. J. Risbeck, N. R. Patel, J. B. Rawlings, Non-linear model predictive control tools for casadi, <https://bitbucket.org/rawlings-group/mpc-tools-casadi>, accessed: 2016-10-10 (2015).
- [31] J. Andersson, A General-Purpose Software Framework for Dynamic Optimization, PhD thesis, Arenberg Doctoral School, KU Leuven, Department of Electrical Engineering (ESAT/SCD) and Optimization in Engineering Center, Kasteelpark Arenberg 10, 3001-Heverlee, Belgium (October 2013).