Linear Moving Horizon Estimation With Pre-estimating Observer

Dan Sui, Tor Arne Johansen and Le Feng

Abstract

In this paper, a moving horizon estimation (MHE) strategy for detectable linear systems is proposed. Like the idea of 'pre-stabilizing' model predictive control, the states are estimated by a forward simulation with a pre-estimating observer in the MHE formulation. Compared with standard linear MHE approaches, it has more degrees of freedom to optimize the noise filtering. Tuning parameters are chosen to minimize the effects of measurement noise and model errors, which is implemented by finding tightest estimation error bounds. The performance of the proposed observer is demonstrated on one linear discrete-time example.

Index Terms

Moving horizon estimation; Luenberger observer; Weight parameters.

I. INTRODUCTION

Deterministic approaches to state estimation are commonly based on Luenberger's design [1], [2]. An alternative method is to use a Kalman filter [3], [4], which is the optimal estimator when the white noise has a Gaussian probability distribution. Unfortunately, the Kalman filter is based on various stochastic assumptions on noise and disturbances that are rarely met in practice, and in combination with model uncertainty. This may lead to inferior performance of the Kalman filter.

Another kind of state estimators is the class of optimization-based state estimators to which moving horizon state estimators belong [5], [6], [7]. The idea of MHE is to estimate the current states by solving a least squares optimization problem, which penalizes the deviation between the measurements and predicted outputs and possibly the distance from the estimated state and a priori information state. The basic strategy is to estimate the state using a moving window of data, such that the size of the estimation is fixed by looking at only a subset of the available information [8]. Since MHE is based on a batch of the most recent information, it can provide a high degree of robustness in the presence of modeling uncertainties [9]. Some empirical studies [10] show that MHE can perform better than the Kalman filter in terms of accuracy and robustness.

In this paper, an improved linear MHE approach based on the work [9] is proposed. In [9], the state dynamics used in the MHE formulation equals the open loop dynamics, which may result in a large propagation of estimation error, especially when the estimation horizon N is large or the system is unstable. Our proposed MHE is similar to the so-called 'pre-stabilizing' model predictive control [11], [12], where the control sequence is parameterized as perturbations to a given pre-stabilizing feedback gain and the states are estimated by a forward simulation with a pre-estimator before optimized in MHE. Its motivation is to reduce the accumulation of the estimation error by using output injection feedback for stabilization and shaping the dynamics. Such a pre-estimator can be chosen as Luenberger observer or others. Another contribution is that detectable rather than observable systems are considered. A crucial issue of MHE is the choice of weight parameters in the cost function. Weight parameters should be considered as the tradeoff between minimizing the effects of measurement noise, model errors, initial state estimation error and possibly other uncertainty. Performance of MHE as a function of weight parameters is generally not easy to analyze since it depends on many factors. Some experiences about choosing weight parameters are discussed in [13], [14]. In this paper, instead of using scalar weight parameters, a weight matrix is introduced into the cost function. Furthermore, weight parameters are determined by minimizing the upper bound of the estimation errors, which is useful especially when there is no statistical information about the system disturbances and measurement noise.

The outline of the paper is as follows: After the introduction, the linear MHE is introduced in Section 2. Following it, its convergence is shown in Section 3. Choosing weight parameters is presented in Section 4, and the performance of the proposed observer is demonstrated in Section 5. Finally, final discussion and conclusions are given in Section 6.

The following notation and nomenclature is used. Positive definite (semi-definite) square matrix A is denoted by $A \succ 0$ ($A \succeq 0$). $\|\cdot\|$ is the Euclidean norm. Let $\rho(A)$ denote spectral radius of a square matrix A respectively; $\underline{\lambda}(A)$ is the smallest eigenvalue of A and $\det(A)$ is its determinate. For a generic matrix M, $||M||_{min} = \sqrt{\underline{\lambda}(M^T M)}$. $I_r \in R^{r\times r}$ is an identity matrix.

II. ESTIMATOR FORMULATION

The following discrete-time, linear time-invariant system is considered,

$$x_{t+1} = Ax_t + Bu_t + \xi_t, \tag{1}$$

$$y_t = Cx_t + \eta_t, \tag{2}$$

D.Sui, T. A. Johansen and L.Feng are with Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway.

where $x_t \in R^{n_x}, u_t \in R^{n_u}$ and $y_t \in R^{n_y}$ are the state, input and the measurement, respectively, $\xi_t \in \Xi \subset R^{n_x}$ is the unmodelled system disturbance, $\eta_t \in \Sigma \subset R^{n_y}$ is the measurement noise and t is the discrete time index.

Moving horizon estimation recursively estimates the state by considering a finite window of data. The problem consists in estimating, at any time $t = N, N+1, \ldots$, the state vectors $x_{t-N}, \ldots, x_t$, on the basis of the a priori estimator $\overline{x}_{t-N,t}$ and the information vector defined as $\mathscr{I}_t = col(y_{t-N}, \ldots, y_t, u_{t-N}, \ldots, u_{t-1})$, where $N+1$ is the window length or horizon. At any stage $t = N, \quad N+1, \ldots$, the linear MHE problem in [9] is formulated as

$$J(\hat{x}_{t-N,t}; \overline{x}_{t-N,t}, \mathscr{I}_t) = \|y^t_{t-N} - \hat{y}^{t,t}_{t-N,t}\|^2 + \alpha \|\hat{x}_{t-N,t} - \overline{x}_{t-N,t}\|^2 \tag{3a}$$

subject to

$$\hat{x}_{i+1,t} = A\hat{x}_{i,t} + Bu_i, \quad i = t-N, \ldots, t-1, \tag{3b}$$

$$\hat{y}_{i,t} = C\hat{x}_{i,t}, \quad i = t-N, \ldots, t, \tag{3c}$$

with $\alpha \geq 0, y^t_{t-N} = col(y_{t-N}, \ldots, y_t)$ and $\hat{y}^{t,t}_{t-N,t} = col(\hat{y}_{t-N,t}, \ldots, \hat{y}_{t,t})$. In (3b), $\hat{x}_{i,t}$ is obtained by a forward simulation, which may accumulate the estimation error, especially when N is large and (A, B) has unstable dynamics. In this paper we improve the MHE strategy (3) by introducing a Luenberger observer into (3b) to pre-estimate $\hat{x}_{i,t}$, since the injection term will reduce the effect of model uncertainty in the a priori estimate and thereby contribute to improve the accuracy. The proposed MHE problem is formulated, as follows,

$$J(\hat{x}_{t-N,t}; \overline{x}_{t-N,t}, \mathscr{I}_t) = \|W(y^t_{t-N} - \hat{y}^{t,t}_{t-N,t})\|^2 + \alpha \|\hat{x}_{t-N,t} - \overline{x}_{t-N,t}\|^2 \tag{4a}$$

subject to

$$\hat{x}_{i+1,t} = A\hat{x}_{i,t} + Bu_i + L(y_i - \hat{y}_{i,t}), \quad i = t-N, \ldots, t-1, \tag{4b}$$

$$\hat{y}_{i,t} = C\hat{x}_{i,t}, \quad i = t-N, \ldots, t, \tag{4c}$$

where $W \in R^{n_x \times (N+1)n_y}$ is a weight matrix and $L \in R^{n_x \times n_y}$ is a gain matrix. The analysis of the proposed MHE requires system (1)-(2) satisfies the following assumptions,

- (A1) Ξ and Σ are compact sets.
- (A2) System (1)-(2) is detectable [15]. Without loss of generality we assume that system (1)-(2) is with $A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$, $C = (0, C_2)$, where the pair (A_{22}, C_2) is observable and $A_{11} \in R^{n_{x1} \times n_{x1}}$ is Hurwitz.
- (A3) L is chosen such that $\Phi = A - LC$ is Hurwitz.

Uniform observability (in some form, see also [16], [17], [18]) is assumed for stability or convergence proofs in the above mentioned references. Uniform observability ensures that the system and data are such that the state estimate solution exists, is unique and depends continuously on the measurement data. However, uniform observability is a restrictive assumption that is likely not to hold in certain interesting and important state estimation applications. This is in particular true for combined state and parameter estimation problems where convergence of the parameter estimates in an augmented state space model will depend on the information contents in the data, typically formulated as a condition for persistently exciting (PE) input data appearing in adaptive control and estimation, e.g. [19], [20], [21]. Therefore, in this paper a detectable system is considered.

The optimal solution of (4) is defined by $\hat{x}^o_{t-N,t}$ and it yields the sequence of the state estimates $\hat{x}^o_{i,t}, i = t-N, \ldots, t$. It is assumed that the a priori estimator is determined from $\hat{x}^o_{t-N-1,t-1}$, that is

$$\overline{x}_{t-N,t} = A\hat{x}^o_{t-N-1,t-1} + Bu_{t-N-1} + L(y_{t-N-1} - \hat{y}^o_{t-N-1,t-1}), \tag{5a}$$

$$\hat{y}^o_{t-N-1,t-1} = C\hat{x}^o_{t-N-1,t-1}. \tag{5b}$$

Then $\overline{x}_{t-N,t}$ can be written as $\overline{x}_{t-N,t} = \Phi\hat{x}^o_{t-N-1,t-1} + Bu_{t-N-1} + Ly_{t-N-1}$. Let $J^o_t = J(\hat{x}^o_{t-N,t}; \overline{x}_{t-N,t}, \mathscr{I}_t)$. The estimation error is defined as

$$e_{t-N} = x_{t-N} - \hat{x}^o_{t-N,t}. \tag{6}$$

In the proposed MHE formulation (4), we perform a forward simulation with a Luenberger observer instead of performing a open-loop forward simulation, which may reduce the estimation error. In the cost function (4a), a weight matrix W is multiplied to the term $y^t_{t-N} - \hat{y}^{t,t}_{t-N,t}$. The choice of W will be discussed in the next sections. Since the dynamics may be stabilized, the convergence of the proposed MHE only depends on Φ, whatever α is. In contrast, the convergence of the MHE formulation (3) depends on both A and α, see the discussions in [9], which may limit the degree of freedom to choose parameter α. Therefore, with the matrix L , one benefit is to have more degrees of freedom to tune α in order to optimize the noise filtering and obtain the good performance.

III. Stability of linear moving horizon estimation

In order to state the convergence result, we define

$$F_N = \begin{bmatrix} C \\ C\Phi \\ \vdots \\ C\Phi^N \end{bmatrix}, \tilde{F}_N = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^N \end{bmatrix}, G_N = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ CB & 0 & \cdots & 0 \\ C\Phi B & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C\Phi^{N-1}B & C\Phi^{N-2}B & \cdots & CB \end{bmatrix}, \tilde{G}_N = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-1}B & CA^{N-2}B & \cdots & CB \end{bmatrix},$$

$$L_N = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ CL & 0 & \cdots & 0 & 0 \\ C\Phi L & CL & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C\Phi^{N-1}L & C\Phi^{N-2}L & \cdots & CL & 0 \end{bmatrix}, H_N = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C\Phi^{N-1} & C\Phi^{N-2} & \cdots & C \end{bmatrix}, \tilde{H}_N = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-1} & CA^{N-2} & \cdots & C \end{bmatrix},$$

$O_N = F_N^T F_N, Q_N = I_{(N+1)n_y} - L_N.$

Assuming all variables are reasonably scaled, we propose to choose the matrix W such that

$$WF_N = \sqrt{\beta} I_{n_x}, \tag{7}$$

where $\beta > 0$ is a scalar tuning parameter. If the system is observable, it leads to

$$W = \sqrt{\beta} F_N^+, \tag{8}$$

where $F_N^+ = (F_N^T F_N)^{-1} F_N^T$ is the pseudo-inverse. If the system is poorly observable, then W defined by (8) will have large values. A thresholded singular value decomposition inversion (SVD) [22] offers a solution for systems that are detectable or poorly observable. Consider an SVD

$$F_N = USV^T. \tag{9}$$

Any singular value (diagonal elements of the matrix S) that is zero or close to zero indicates that a mode is unobservable or poorly observable. Moreover, the corresponding row of the V matrix will indicate which modes cannot be estimated. In general, we do not want to make the estimation of the unobservable modes depend on the measured data since there is no information, only noise. This is effectively achieved by utilizing the SVD in order to compute a ”robust pseudo-inverse” where the inverse of *small* singular values is set to zero rather than let it grow unbounded. From (8), we have

$$W = \sqrt{\beta} V S_\delta^+ U^T, \tag{10}$$

where the thresholded pseudo-inverse $S_\delta^+ = \mathrm{diag}(0,...,0,1/\sigma_1,...1/\sigma_l)$ where $\sigma_1,...,\sigma_l$ are the singular values larger than some sufficiently small $\delta > 0$ and the zeros correspond to small singular values whose inverse is set to zero [22]. It is necessary to note that too large δ may lead to a bias on estimates also without any noise or disturbances.

Theorem 1: Suppose that assumptions (A1)-(A3) hold and W is chosen according to (10) with some sufficiently small $\delta > 0$, and $\alpha \geq 0, \beta > 0$. Then the error dynamics is given by,

$$e_{t-N} = \bar{A} e_{t-N-1} + \bar{G} \xi_{t-N-1}^{t-1} + \bar{H} \eta_{t-N-1}^{t}, \tag{11}$$

where $\bar{A} = \alpha\Gamma\Phi$, $\bar{G} = [\alpha\Gamma| - \Theta W H_N]$ and $\bar{H} = -[\alpha\Gamma L|\Theta W Q_N]$ with $\Gamma = \begin{bmatrix} \frac{1}{\alpha} I_{n_{x1}} & 0 \\ 0 & \frac{1}{\alpha+\beta} I_{n_{x2}} \end{bmatrix}$ and $\Theta = \begin{bmatrix} 0 & 0 \\ 0 & \frac{\sqrt{\beta}}{\alpha+\beta} I_{n_{x2}} \end{bmatrix}$, $n_{x2} = n_x - n_{x1}$.

Proof: The necessary condition on the minimum of the cost function (4a) is

$$\nabla_{\hat{x}^o_{t-N,t}} J(\hat{x}^o_{t-N,t}; \bar{x}_{t-N,t}, \mathscr{I}_t) = 2\alpha(\hat{x}^o_{t-N,t} - \bar{x}_{t-N,t}) - 2(WF_N)^T W(y_{t-N}^t - F_N \hat{x}^o_{t-N,t} - G_N u_{t-N}^{t-1} - L_N y_{t-N}^t) = 0,$$

for any $t = N, N+1, \ldots$, where $u_{t-N}^{t-1} = col(u_{t-N}, \ldots, u_{t-1})$. It is easy to show that

$$y_{t-N}^t = \tilde{F}_N x_{t-N} + \tilde{G}_N u_{t-N}^{t-1} + \tilde{H}_N \xi_{t-N}^{t-1} + \eta_{t-N}^t,$$

where $\xi_{t-N}^{t-1} = col(\xi_{t-N}, \ldots, \xi_{t-1})$ and $\eta_{t-N}^t = col(\eta_{t-N}, \ldots, \eta_t)$, then we have

$$\begin{aligned} Q_N y_{t-N}^t &= Q_N \tilde{F}_N x_{t-N} + Q_N \tilde{G}_N u_{t-N}^{t-1} + Q_N \tilde{H}_N \xi_{t-N}^{t-1} + Q_N \eta_{t-N}^t \\ &= F_N x_{t-N} + G_N u_{t-N}^{t-1} + H_N \xi_{t-N}^{t-1} + Q_N \eta_{t-N}^t. \end{aligned}$$

From the above, it is obtained that

$$y_{t-N}^t - F_N \hat{x}^o_{t-N,t} - G_N u_{t-N}^{t-1} - L_N y_{t-N}^t = F_N (x_{t-N} - \hat{x}^o_{t-N,t}) + H_N \xi_{t-N}^{t-1} + Q_N \eta_{t-N}^t. \tag{12}$$

Since W is chosen according to (10) with some sufficiently small $\delta > 0$,

$$WF_N = \sqrt{\beta} V S_\delta^+ U^T U S V^T = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\beta} I_{n_{x2}} \end{bmatrix}.$$

Then we have

$$(WF_N)^T W (y_{t-N}^t - F_N \hat{x}_{t-N,t}^o - G_N u_{t-N}^{t-1} - L_N y_{t-N}^t) = (WF_N)^T W F_N (x_{t-N} - \hat{x}_{t-N,t}^o) + (WF_N)^T W H_N \xi_{t-N}^{t-1} + (WF_N)^T W Q_N \eta_{t-N}^t$$
$$= \begin{bmatrix} 0 & 0 \\ 0 & \beta I_{n_{x2}} \end{bmatrix} (x_{t-N} - \hat{x}_{t-N,t}^o) + \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\beta} I_{n_{x2}} \end{bmatrix} W H_N \xi_{t-N}^{t-1} + \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\beta} I_{n_{x2}} \end{bmatrix} W Q_N \eta_{t-N}^t.$$

As $\bar{x}_{t-N,t} = \Phi \hat{x}_{t-N-1,t-1}^o + B u_{t-N-1} + L y_{t-N-1}$, we have

$$\hat{x}_{t-N,t}^o - \bar{x}_{t-N,t} = \hat{x}_{t-N,t}^o - x_{t-N} + x_{t-N} - \bar{x}_{t-N,t} = \hat{x}_{t-N,t}^o - x_{t-N} + \Phi(x_{t-N} - \hat{x}_{t-N-1,t-1}^o) + \xi_{t-N-1} - L\eta_{t-N-1}.$$

It is obtained that

$$\begin{aligned}
&(\alpha I_{nx} + \begin{bmatrix} 0 & 0 \\ 0 & \beta I_{n_{x2}} \end{bmatrix})(x_{t-N} - \hat{x}_{t-N,t}^o) = \alpha \Phi (\hat{x}_{t-N-1,t-1}^o - x_{t-N-1}) + \alpha \xi_{t-N-1} - \alpha L \eta_{t-N-1} \\
&- \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\beta} I_{n_{x2}} \end{bmatrix} (W H_N \xi_{t-N}^{t-1} + W Q_N \eta_{t-N}^t) \\
&\Rightarrow x_{t-N} - \hat{x}_{t-N,t}^o = \alpha \Gamma \Phi (\hat{x}_{t-N-1,t-1}^o - x_{t-N-1}) + \alpha \Gamma \xi_{t-N-1} - \Theta W H_N \xi_{t-N}^{t-1} - \alpha \Gamma L \eta_{t-N-1} - \Theta W Q_N \eta_{t-N}^t.
\end{aligned}$$

By substituting $e_{t-N} = x_{t-N} - \hat{x}_{t-N,t}^o$ into the above equation, we obtain (11). ■

From Theorem 1, we know that the dynamics of the estimation error depends on the choices of L, δ, α, β and N. By designing the injection matrix L, one result is shown below.

Theorem 2: Suppose assumptions (A1)-(A3) hold. If W is chosen according to (10) with some sufficiently small $\delta > 0$, and $\alpha \geq 0, \beta > 0$, then $\rho(\bar{A}) < 1$. Moreover, when $\xi_t = 0, t = 0, 1, \ldots$ and $\eta_t = 0, t = 0, 1, \ldots$, e_t converges to zero.

Proof: Since $\Phi = A - LC$, $A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$ and $C = (0, C_2)$, we have $\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$ with $\Phi_{11} = A_{11}$ and $\Phi_{21} = 0$. Then

$$\bar{A} = \alpha \Gamma \Phi = \begin{bmatrix} I_{n_{x1}} & 0 \\ 0 & \frac{\alpha}{\alpha+\beta} I_{n_{x2}} \end{bmatrix} \begin{bmatrix} A_{11} & \Phi_{12} \\ 0 & \Phi_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & \Phi_{12} \\ 0 & \frac{\alpha}{\alpha+\beta} \Phi_{22} \end{bmatrix}. \tag{13}$$

Since $\alpha \geq 0, \beta > 0$ and $\rho(\Phi) < 1$, we have $\rho(\bar{A}) < 1$. Without disturbances and measurement noises, from (11), it is easy to obtain $e_t \to 0$ when $t \to \infty$. ■

The motivation of choosing the weighting matrix W according to (10) is to avoid F_N involved in the error dynamics (11) by introducing the scaler β to replace the term WF_N. Such a choice makes the analysis of error dynamics very easy. From Theorem 2 no matter what α and β are, by choosing a stable gain L such that $\rho(\Phi) < 1$, $\rho(\bar{A}) < 1$ or the robust exponential convergence of the estimation error is guaranteed. However for MHE problem (3), without disturbances and measurement noises it is shown in [9] that $\|e_{t-N}\| \leq \tilde{q} \|e_{t-N-1}\|$, where $\tilde{q} = \frac{\alpha ||A||}{\alpha + \|\tilde{F}_N\|_{min}^2}$. Then, if $||A|| \leq 1$, one can choose any $\alpha \geq 0$. Instead, when $||A|| > 1$, $\tilde{q} < 1$ holds if $0 \leq \alpha < \frac{\|\tilde{F}_N\|_{min}^2}{||A||-1}$. Therefore, compared with the problem (3), the proposed MHE has more flexibility to tune parameters α, β in order to obtain a better performance. The focus in the following is on the stability and robust convergent properties of the proposed MHE. The behavior of MHE is analyzed when the system (1)-(2) is with bounded disturbances and measurement noises. In the next theorem, it is shown that the solution of the MHE problem provides an estimation error that admits a time-varying upper bound. Let us define

$$r_\xi = \max_{\xi_t \in \Xi} ||\xi_t||, \; r_\eta = \max_{\eta_t \in \Sigma} ||\eta_t||, \quad t = 0, 1, \ldots.$$

Theorem 3: Suppose assumptions (A1)-(A3) hold and W is chosen according to (10) with some sufficiently small $\delta > 0$. Then the norm of the estimation error is bounded as

$$||e_{t-N}|| \leq \zeta_{t-N}$$

where $\{\zeta_t\}$ is a sequence generated by

$$\zeta_0 = b_0, \tag{14a}$$
$$\zeta_t = a \zeta_{t-1} + b, \quad t = 1, 2, \ldots \tag{14b}$$

with

$$
\begin{aligned}
a &= ||\Phi|| \\
b &= r_\xi + \frac{\sqrt{\beta N}}{\alpha+\beta} fhr_\xi + \ell r_\eta + \frac{\sqrt{\beta(N+1)}}{\alpha+\beta} fqr_\eta, \\
b_0 &= ||x_0 - \bar{x}_{0,N}|| + \frac{\sqrt{\beta N}}{\alpha+\beta} fhr_\xi + \frac{\sqrt{\beta(N+1)}}{\alpha+\beta} fqr_\eta,
\end{aligned}
$$

and $f = ||W||, h = ||H_N||, \ell = ||L||, q = ||Q_N||$. Furthermore if α, β and L are selected such that $a < 1$, then the bounding sequence $\{\zeta_t\}$ has the following properties:
(a) $\{\zeta_t\}$ converges exponentially to the asymptotic value $e_\infty(\alpha,\beta,L,N) = \frac{b}{1-a}$;
(b) If $\zeta_t > e_\infty(\alpha,\beta,L,N)$, then $\zeta_{t+1} < \zeta_t, \quad t = 0,1,\ldots$.

Proof: The error dynamics is shown in (11). The asymptotic upper bound can be easily derived when $a < 1$. And

$$\zeta_t = a^t \zeta_0 + b \sum_{j=0}^{t-1} a^j,$$

which tends to $\frac{b}{1-a}$ as $t \to \infty$. Then property (a) is proven. Due to the fact that $\zeta_t - \zeta_{t+1} = (1-a)\zeta_t - b$, and $\zeta_t > \frac{b}{1-a}$, we have $\zeta_t - \zeta_{t+1} > (1-a)\frac{b}{1-a} - b = 0$. Property (b) is shown. ■

In the literature, there is a solid analytical method to select the matrix L, see [1], [2]. It is important to choose the parameters α and β. From Theorem 3, when $a < 1$, the asymptotic estimation error is,

$$e_\infty(\alpha,\beta,L,N) = \frac{b}{1-a}. \tag{15}$$

It is clear that increased a leads to the increased filtering. Unfortunately, the behavior of (15) as a function of α, β and N is generally not easy to analyze and provide more detailed hint how to tune α and β, as it depends on many parameters. In the next section, a computational approach to tune α and β will be introduced.

IV. Tuning parameters

The motivation for choosing $\alpha > 0$ is twofold. First, it will lead to filtering being an integral part of the state estimator, see [9] for a discussion of the filtering effect. Second, it will allow us to more systematically handle several cases. When $\alpha = 0$, the estimator results in a dead-beat observer. When $\beta = 0$ and $\rho(\Phi) < 1$, the proposed MHE estimator essentially becomes a Luenberger observer.

In this paper, the weight parameters are chosen based on the following principles

- $\rho(\bar{A}) < 1$;
- The upper bound of e_t is as small as possible.

Theorem 3 provides one way to estimate the upper bound on the estimation error, which needs the condition $a < 1$. In the literature, there exist other approaches to compute or approximate such bound. For example, in [9], upper bounds based on pure triangular inequalities are derived. Less conservative bounds are given in [13]. The error dynamics (11) can be rewritten as

$$e_t = \bar{A} e_{t-1} + \bar{E} \mathbf{w}_{t-1}, \tag{16}$$

where $\mathbf{w}_{t-1} = col(\xi_{t-N-1},\ldots,\xi_{t-1},\eta_{t-N-1},\ldots,\eta_t)$ and $\bar{E} = [\bar{G},\bar{H}]$. Suppose noise and disturbances further satisfy the following assumption

- (A4) $\mathbf{w}$ in (16) belongs to an ellipsoidal compact set $\varepsilon_Q := \{\mathbf{w} : \mathbf{w}^T Q \mathbf{w} < 1\}$ with $Q \succ 0$.

Theorem 4: [13] Suppose assumptions (A1)-(A4) hold and W is chosen according to (10) with some sufficiently small $\delta > 0$, and $\alpha \geq 0, \beta > 0$. The following facts are equivalent:
(1) The system (16) is strictly quadratically bounded with a common Lyapunov matrix $P \succ 0$ for all $\mathbf{w} \in \varepsilon_Q$.
(2) The ellipsoid $\varepsilon_P := \{e : e^T P e < 1\}$ is a positively invariant set for the system (16) for all $\mathbf{w} \in \varepsilon_Q$.
(3) There exists $\mu > 0$ such that

$$\begin{bmatrix} \bar{A}^T P \bar{A} - P + \mu P & \bar{A}^T P \bar{E} \\ \bar{E}^T P \bar{A} & \bar{E}^T P \bar{E} - \mu Q \end{bmatrix} \preceq 0. \tag{17}$$

■

From Theorem 4, to obtain the smallest bound on e_t in terms of the volume of ε_P is the same as to find the matrix $P \succ 0$ that minimizes $-\ln\det(P)$ [23]. Therefore, one can obtain optimal α and β by solving the following minimization problem. It is sufficient to fix β and optimization α, or vice versa, since only $\frac{W}{\sqrt{\alpha}}$ is important in (4a).

Inner-loop minimization:

$$V(P) = \min_P (-\ln\det(P)) \tag{18a}$$

subject to

$$\begin{bmatrix} -(1-\mu)P & 0 & \bar{A}^T P \\ 0 & -\mu Q & \bar{E}^T P \\ P\bar{A} & P\bar{E} & -P \end{bmatrix} \preceq 0, \tag{18b}$$

$$P \succ 0, \tag{18c}$$

where (18b) is equivalent to (17), see [24].

Outer-loop minimization:

$$\min_{\mu,\alpha,\beta} V(P) \tag{19a}$$

subject to

$$\mu > 0, \tag{19b}$$

$$\alpha \geq 0, \tag{19c}$$

$$\beta > 0. \tag{19d}$$

Please note for any given α, β and μ the inner-loop minimization is a linear matrix inequality (LMI) [25] problem, which can be efficiently solved with some existing toolboxes (LMIlab, SeDuMi [26], etc.). The outer-loop minimization is actually nonlinear over three scalers μ, α and β. It is a non-convex optimization problem which generally requires good initializations. In the example we use TOMLAB 'npsol' algorithm using a BFGS Quasi-Newton method. Assuming optimality of the solution we can define $(P^o, \mu^o, \alpha^o, \beta^o) = \text{argmin}(-\ln\det(P))$ under the constraints (18b)-(18c) and (19b)-(19d).

V. Example

The computational advantages will be shown based on a highly nonlinear model of a continuous stirred tank reactor (CSTR), see [27]. Assuming constant liquid volume, the CSTR for an exothermic, irreversible reaction, $A \to B$, is described by the following dynamic model based on a component balance for reactant A and an energy balance:

$$\begin{aligned} \dot{C}_A &= \frac{q}{V}(C_{Af} - C_A) - k_0 exp(-\frac{E}{RT})C_A \\ \dot{T} &= \frac{q}{V}(T_f - T) + \frac{-\Delta H}{\rho C_p} k_0 exp(-\frac{E}{RT})C_A + \frac{UA}{V\rho C_p}(T_c - T) \end{aligned} \tag{20}$$

where C_A is the concentration of A in the reactor, T is the reactor temperature, and T_c is the temperature of the coolant stream. The objective is to regulate C_A and T by manipulating T_c. The nominal operating conditions, which correspond to an unstable equilibrium $C_A^{eq} = 0.5mol/l$, $T^{eq} = 350K$, $T_c^{eq} = 300K$ are: $q = 100l/min$, $T_f = 350K$, $V = 100l$, $\rho = 1000g/l$, $C_p = 0.239J/gK$, $\Delta H = -5 \times 10^4 J/mol$, $E/R = 8750K$, $k_0 = 7.2 \times 10^{10} min^{-1}$, $UA = 5 \times 10^4 J/minK$. Using a sampling time of $t_s = 0.1$min and introducing deviation variables (from the corresponding steady state) a linearized model is as follows,

$$x_{t+1} = \begin{bmatrix} 0.9384 & -0.0011 \\ 6.5063 & 1.1372 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 0.0675 \end{bmatrix} u_t + \xi_t, \quad y_t = [0 \quad 1]x_t + \eta_t, \tag{21}$$

in which x_1 and x_2 represent the reactant concentration and the reactor temperature, respectively; u represents the coolant temperature. Suppose ξ and η are independent uniformly distributed noise and disturbances with $\xi \in [-0.05, 0.05]$ and $\eta \in [-0.05, 0.05]$. Let us assume x_0 and $\bar{x_0}$ to be independent uniformly distributed random variables with $x_0 \in [(-0.5, -20), (0.5, 20)]$ and $\bar{x_0} \in [(-0.5, -20), (0.5, 20)]$. In the simulation, we choose randomly 1000 initial states that are used to evaluate the performance for the different cases. Choose $u = Kx$ where $K = [-101.1489, -4.7982]$, which is computed as the solution of the LQ gain with (Q, R) being $(I_2, 0.1)$. Since initial guess of states impacts on transient state error, to fairly compare, the steady-state errors are given in the example. Let us consider the performance indexes given by the root mean square error (RMSE):

$$RMSE = (\sum_{t=t_P}^{t_S} \frac{(||e_t||^2}{t_S - t_P})^{1/2},$$

where $||e_t||$ is the norm of the estimation error at time t, t_S is the length of each simulation run. We choose $t_S = 60$ and $t_P = 10$.

Consider the following cases:

- Case 1. $N = 4$. Choose $L = [0.1486, 2.1754]^T$ such that the eigenvalues of Φ are $(0, -0.1)$ and W is chosen by (10). The weight parameters α and β are obtained by solving problem (18).
- Case 2. $N = 4$. Choose $L = [0, 0]^T$ and W is chosen by (10). The weight parameters α and β are obtained by solving problem (18).
- Case 3. $N = 4$. Choose $L = [0, 0]^T$ and $W = I_{(N+1)n_y}$. The weight parameter α is obtained by solving problem (18).
- Case 4. $N = 10$. L and W are the same as Case 1. The weight parameters α and β are obtained by solving problem (18).

- Case 5. $N = 10$. L and W are the same as Case 2. The weight parameters α and β are obtained by solving problem (18).
- Case 6. $N = 10$. L and W are the same as Case 3. The weight parameters α and β are obtained by solving problem (18).
- Case 7. Choose $L = [0.1486, 2.1754]^T$ and $\beta = 0$.
- Case 8. Choose the steady-state Kalman-filter. Assume that the noise covariance is 0.0167 and disturbance covariance is 0.0167, then $L = [-0.0012, 0.2101]^T$.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Average RMSE	0.0117	0.0914	0.0921	0.0126	0.3219	0.3219	0.1624	0.0562

TABLE I
THE PERFORMANCES OF CASES 1-8.

Table I presents the performances of the above cases. Compared with Case 1, 2 and 3, we know that the performance of Case 1 with $L \neq \mathbf{0}$ seems to be best. When N is increasing, see Case 4, the performance of the proposed MHE may be improved due to more information data involved. However, in the case of $L = \mathbf{0}$, since a large N may result in a large propagation error due to the open-loop unstable dynamics, the performance may become worse in Case 5 (6) compared with Case 2 (3). Case 7 is just a special case of the proposed MHE when choosing $\beta = 0$ which corresponds to a Luenberger observer. Therefore, compared with Luenberger observer, our approach has more freedom to choose parameters α, β to obtain better performance.

VI. CONCLUSION

In this paper, a moving horizon estimation for linear systems is proposed. It leads to a good performance especially when the estimation horizon N is large or the control system has slow or unstable dynamics. The proposed MHE can handle the case when the system is poorly observable or only detectable by using robust pseudo-inverse to choose the weight matrix W. By introducing a feedback error into the estimated state, the estimator can avoid an unnecessary propagation error and the effect of disturbance, noises and model errors are reduced. In general, tuning the weight parameters is a key point in the MHE processing. We proposed an approach to choose α and β by minimizing the upper bounds of errors. From the simulation result, it is easy to know that the proposed MHE has more flexibility to choose weight parameters and has more freedom to obtain a good behavior.

ACKNOWLEDGEMENTS

This work is supported by the Research Council of Norway Strategic University Program on Computational Method in Nonlinear Motion Control.

REFERENCES

[1] D. G. Luenberger. Observers for multivariable systems. *IEEE Transactions on Automatic Control*, pages 190–197, 1966.
[2] D. G. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control*, pages 596–602, 1971.
[3] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASMECJournal of Basic Engineering*, 82:35–45, 1960.
[4] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* Wiley-Interscience, 2006.
[5] C. V. Rao, J. B. Rawlings, and D. Q. Mayne. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximation. *IEEE Transactions Automatic Control*, 48:246–258, 2003.
[6] P. E. Moraal and J. W. Grizzle. Observer design for nonlinear systems with discrete-time measurement. *IEEE Transactions Automatic Control*, 40:395–404, 1995.
[7] A. Alessandri, M. Baglietto, and G. Battistelli. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*, 44:1753–1765, 2008.
[8] C. V. Rao, J. B. Rawlings, and J. H. Lee. Constrained linear state estimation: a moving horizon approach. *Automatica*, 37:1619–1628, 2001.
[9] A. Alessandri, M. Baglietto, and G. Battistelli. Receding horizon estimation for discrete time linear systems. *IEEE Transaction on Automatic Control*, 48, 2003.
[10] E. L. Haseltine and J. B. Rawlings. Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Ind. Eng. Chem. Res*, 44:2451–2460, 2005.
[11] J. A. Rossiter, B. Kouvaritakis, and M. J. Rice. A numerically robust state-space approach to stable predictive control strategies. *Automatica*, 34:65–73, 1998.
[12] B. Kouvaritakis, J. A. Rossiter, and J. Schuurmans. Efficient robust predictive control. *IEEE Transactions on Automatic Control*, 45(8):1545–1549, 2000.
[13] A. Alessandri, M. Baglietto, and G. Battistelli. On estimation error bounds for receding horizon filters using quadratic boundedness. *IEEE Transaction on Automatic Control*, 49, 2004.
[14] A. Alessandri, M. Baglietto, T. Parisini, and R. Zoppoli. A neural state estimator with bounded errors for nonlinear systems. *IEEE Transaction on Automatic Control*, 44:2028–2042, 1999.
[15] C. T. Chen. *Linear System Theory and Design.* Oxford University Press, 1998.
[16] T. Raff, C. Ebenbauer, R. Findeisen, and F. Allgöwer. Remarks on moving horizon state estimation with guaranteed convergence. In T. Meurer, K. Graichen, and E. D. Gilles, editors, *Control and Obsever Design for Nonlinear Finite and Infinite Dimensional Systems*, pages 67–80. Springer-Verlag, Berlin, 2005.
[17] M. Alamir. Optimization based non-linear observers revisited. *Int. J. Control*, 72:1204–1217, 1999.
[18] E. D. Sontag. A concept of local observability. *Systems and Control Letters*, 5:41–47, 1984.
[19] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Inc., Englewood Cliffs, NJ., 1999.
[20] G. Kreisselmeier. Adaptive observers with exponential rate of convergence. *IEEE Trans. Automatic Control*, 22:2–8, 1977.

[21] M. Krstic, I. Kanellakopoulos, and P. Kokotovic. *Nonlinear and Adaptive Control Design*. Wiley and Sons, 1995.
[22] G. H. Golub and C. F. van Loan. *Matrix computations*. Oxford University Press, 1983.
[23] Johan Löfberg. Yalmip: A matlab interface to sp, maxdet and socp. Technical Report LiTH-ISY-R-2328, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, January 2001.
[24] Fuzhen Zhang, editor. *The Schur Complement and Its Applications*. Number 4 in Numerical Methods and Algorithms. Springer, Boston, MA, United States, April 2005.
[25] S. P. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. Linear matrix inequalities in system and control theory. *SIAM*, 1998.
[26] Jos F. Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, pages 625–653, 1999.
[27] L. Magni, G. D. Nicolao, L. Magnani, and R. Scattolini. A stablilizing model-based predictive control algorithm for nonlinear systems. *Automatica*, 31:1351–1362, 2001.