International Journal of Control
Vol. 00, No. 00, Month 200x, 1–21

RESEARCH ARTICLE

Moving Horizon Observer with Regularization for Detectable Systems without Persistence of Excitation

Sui Dan and Tor A. Johansen*
Department of Engineering Cybernetics, Norwegian University of Science and Technology, Trondheim, Norway.
(*Received 00 Month 200x; final version received 00 Month 200x*)

A constrained moving horizon observer is developed for nonlinear discrete-time systems. The algorithm is proved to converge exponentially under a detectability assumption and the data being exciting at all time. However, in many practical estimation problems, such as combined state and parameter estimation, the data may not be exciting for every period of time. The algorithm therefore has regularization mechanisms to ensure robustness and graceful degradation of performance in cases when the data are not exciting. This includes the use of a priori estimates in the moving horizon cost function, and the use of thresholded singular value decomposition to avoid ill-conditioned or ill-posed inversion of the associated nonlinear algebraic equations that define the moving horizon cost function. The latter regularization relies on monitoring of the rank of an estimate of a Hessian-like matrix and conditions for uniform exponential convergence are given. The method is in particular useful with augmented state space models corresponding to mixed state and parameter estimation problems, or dynamics that are not asymptotically stable, as illustrated with two simulation examples.

Keywords: State Estimation; Nonlinear Systems; Regularization; Detectability;

1 Introduction

Consider the state estimation problem of nonlinear discrete-time systems. A least-squares optimal state estimation problem can be formulated by minimizing a properly weighted least-squares criterion defined on the full horizon of the data history, subject to the nonlinear model equations, Moraal and Grizzle (1995), Rao et al. (2003). This is, however, impractical as infinite memory and processing will be needed as the amount of data grows unbounded with time. Alternatively, a well known sub-optimal estimator is given by an Extended Kalman Filter (EKF) which approximates this least-squares problem and defines a finite memory recursive algorithm suited for real-time implementation, where only the last measurement is used to update the state estimate, based on the past history being approximately summarized by estimates of the state and the error covariance matrix, Gelb (2002). Unfortunately, the EKF is based on various stochastic assumptions on noise and disturbances that are rarely met in practice, and in combination with nonlinearities and model uncertainty, this may lead to unacceptable performance of the EKF. A possible better use of the dynamic model and past history when updating the state estimate is made by a Moving Horizon State Estimator (MHE) that makes use of a finite memory moving window of both current and historical measurement data in the least-squares criterion, possibly in addition to a state estimate and covariance matrix estimate to set the initial conditions at the beginning of the data window, see Rao et al. (2003), Moraal and Grizzle (1995), Alessandri et al. (1999, 2008) for different formulation relying on somewhat different assumptions. Such an MHE can also be considered a sub-optimal approximation to an estimator that uses the full history of past data, and some empirical studies, Haseltine and Rawlings

*Corresponding author. Email: tor.arne.johansen@itk.ntnu.no

ISSN: 0020-7179 print/ISSN 1366-5820 online

DOI: 10.1080/0020717YYxxxxxxxx
http://www.informaworld.com

(2005) show that the MHE can perform better than the EKF in terms of accuracy and robustness.

A direct approach to the deterministic discrete-time nonlinear MHE problem is to view the problem as one of inverting a sequence of nonlinear algebraic equations defined from the state update and measurement equations, and some moving time horizon. Such discrete-time observers are formulated in the context of numerical nonlinear optimization and analyzed with respect to stability in Moraal and Grizzle (1995), Alessandri et al. (2008). Some earlier contributions based on similar ideas are given in Glad (1983), Zimmer (1994), while Biyik and Arcak (2006) provides results on how to use a continuous time model in the discrete time design. As pointed out in Grossman (1999), the dead beat type of design philosophy (Moraal and Grizzle (1995)) does not explicitly take into account robustness to noise, and some modifications are required as proposed in Grossman (1999). It should be mentioned that common to all methods is the use of numerical methods subject to the underlying assumption that local minima and multiple solutions may restrict convergence properties to be only local.

Uniform observability (in some form, see also Raff et al. (2005), Alamir (1999)) is assumed for stability or convergence proofs in the above mentioned references, including the EKF, Reif et al. (1998), Reif and Unbehauen (1999). Uniform observability means that the system and data are such that the problem of inverting the nonlinear algebraic equations is well-posed in the sense of Tikhonov and Arsenin (1977), i.e. that the state estimate solution exists, is unique and depends continuously on the measurement data. In the context of optimization this is commonly referred to as stability of the solution, that can be guaranteed under certain regularity and rank conditions, Fiacco (1983). This robustness is essential in any practical application since otherwise the estimates will be impossible to compute, and will be divergent or highly sensitive to imperfections such as numerical round-off errors, quantization and measurement noise. 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 the state space model is augmented with the unknown parameters, Gelb (2002), and convergence of the parameter estimates 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. Krstic et al. (1995), or boundedness of the EKF covariance matrix estimate, Reif et al. (1998), Reif and Unbehauen (1999). In many practical applications the data will be sufficiently exciting for significant periods of time, but may in some time intervals contain insufficient excitation and information. It should also be noted that with some exceptions (e.g. Panteley et al. (2001), Sedoglavic (2002)), both uniform observability and PE conditions are difficult to verify a priori.

In this paper we consider strongly detectable systems Moraal and Grizzle (1995), and the objective and novel contribution of the present work is to provide and study an MHE method based on Alessandri et al. (2008) and others with alternative weighting and regularization to achieve satisfactory practical performance also when the condition of uniform observability is violated due to temporarily lack of persistence of excitation, or the system not being observable. The relaxation to detectability was envisioned in Alessandri et al. (2008), although no proofs of the convergence were given. Following the spirit of Moraal and Grizzle (1995) we introduce practical regularization mechanisms that monitor and estimate the information contents and degree of excitation in the data, and take corresponding action by adaptively weighting the measured data and a priori estimates from the dynamic model. Although the MHE formulation based on Alessandri et al. (2008) does not rely on an explicit uncertainty estimate in terms of a covariance matrix estimate (unlike formulations that may apply an arrival cost estimate, Rao et al. (2003)), the monitoring of persistent excitation in the moving horizon nonlinear observer relies on a related Hessian matrix estimate. This makes the approach similar in spirit to well known modifications of the EKF and Recursive Least Squares estimation methods that rely on monitoring and resetting of the covariance matrix estimate (Salgado et al. (1988)), directional forgetting (Kulhavy (1987), Fortescue et al. (1981), Cao and Schwartz (1999), Campi (1994), Hovd and Bitmead (2007)) and using singular value decomposition for numerically robust matrix inversion.

The outline of the paper is as follows: After the introduction, a description of the nonlinear moving horizon estimation problem and the relevant assumptions are given in Section 2, together with an analysis of its convergence under strong observability and with informative data. Section 4 extends the observer to have graceful degradation and practical performance also for the case when the data are not informative and other assumptions are violated, which is followed by two numerical examples presented in Section 4. Final discussion and conclusions are given in Section 5.

Preliminary results related to this work was published in Sui and Johansen (2010).

The following notation and nomenclature is used. For a vector $x \in \mathbb{R}^n$, let $||x|| = \sqrt{x^T x}$ denote the Euclidean norm. Recall that the induced matrix norm $||M||$ equals the largest singular value of M. For two vectors $x \in R^n$ and $y \in R^m$ we let $\mathrm{col}(x,y)$ denote the column vector in R^{n+m} where x and y are stacked into a single column. The Moore-Penrose pseudo-inverse, Golub and van Loan (1983), of a matrix M is denoted M^+ and we recall that for a matrix M of full rank it is given by $M^+ = (M^T M)^{-1} M^T$ while in general it is defined as $M^+ = VS^+U^T$ where $M = USV^T$ is a singular value decomposition where S is a diagonal matrix with the singular values $\sigma_1, ..., \sigma_n$ on the diagonal, and S^+ is the diagonal matrix $S = \mathrm{diag}(1/\sigma_1, ..., 1/\sigma_r, 0, ..., 0)$ where $r \leq n$ of the singular values are non-zero. The composition of two functions f and g is written $f \circ g(x) = f(g(x))$. Finally, a function $\varphi : R^+ \to R$ is called a K-function if $\varphi(0) = 0$ and it is strictly increasing.

2 Nonlinear MHE problem formulation

Consider the following discrete-time nonlinear system:

$$x_{t+1} = f(x_t, u_t) \tag{1a}$$

$$y_t = h(x_t, u_t), \tag{1b}$$

where $x_t \in \mathbb{X} \subseteq R^{n_x}$, $u_t \in \mathbb{U} \subseteq R^{n_u}$ and $y_t \in R^{n_y}$ are respectively the state, input and measurement vectors, and t is the discrete time index. The sets $\mathbb{X}$ and $\mathbb{U}$ are assumed to be convex and compact. The $N+1$ consecutive measurements of outputs and inputs until time t are denoted as $Y_t = \mathrm{col}(y_{t-N}, y_{t-N+1}, \cdots, y_t)$ and $U_t = \mathrm{col}(u_{t-N}, u_{t-N+1}, \cdots, u_t)$. To express Y_t as a function of x_{t-N} and U_t, denote $f^{u_t}(x_t) = f(x_t, u_t)$ and $h^{u_t}(x_t) = h(x_t, u_t)$, and note from (1b) that the following algebraic map can be formulated, Moraal and Grizzle (1995):

$$Y_t = H(x_{t-N}, U_t) = H_t(x_{t-N}) = \begin{bmatrix} h^{u_{t-N}}(x_{t-N}) \\ h^{u_{t-N+1}} \circ f^{u_{t-N}}(x_{t-N}) \\ \vdots \\ h^{u_t} \circ f^{u_{t-1}} \circ \cdots \circ f^{u_{t-N}}(x_{t-N}) \end{bmatrix}. \tag{2}$$

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

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

Definition 2a. The input $U_t \in \mathbb{U}^{N+1}$ is said to be *N-exciting* for the N-observable system (1) at time t if there exists a K-function φ_t that for all $x_1, x_2 \in \mathbb{X}$ satisfies

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

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

$$H(x_1,U_t) - H(x_2,U_t) = \Phi_t(x_1,x_2)(x_1 - x_2), \tag{3}$$

where

$$\Phi_t(x_1,x_2) = \int_0^1 \frac{\partial}{\partial x} H((1-s)x_2 + sx_1, U_t) ds. \tag{4}$$

Like in the linear case, an observability rank condition can be formulated (see also Moraal and Grizzle (1995), Alessandri et al. (2008), Fiacco (1983) and others for similar results):

Lemma 2.1: *If $\mathbb{X}$ and $\mathbb{U}$ are compact and convex sets, the functions f and h are twice differentiable on $\mathbb{X} \times \mathbb{U}$ and the Jacobian matrix $\frac{\partial H}{\partial x}(x,U_t)$ has full rank (equal to n_x) for all $x \in \mathbb{X}$ and some $U_t \in \mathbb{U}^{N+1}$, then the system is N-observable and the input U_t is N-exciting for the system (1) at time t.*

Proof. Due to the observability rank condition being satisfied, $\Phi_t^T(\cdot)\Phi_t(\cdot) > 0$ and the system of nonlinear algebraic equations (3) can be inverted as follows:

$$x_1 - x_2 = \Phi_t^+(x_1,x_2)\big(H(x_1,U_t) - H(x_2,U_t)\big),$$
$$\Rightarrow \frac{1}{\pi_t^2(x_1,x_2)} \|x_1 - x_2\|^2 \leq \|H(x_1,U_t) - H(x_2,U_t)\|^2,$$

where $\pi_t(x_1,x_2) = \|\Phi_t^+(x_1,x_2)\|$. This proves that the conditions in Definitions 1 and 2a hold with $\varphi(s) = s/\overline{p}^2$ where $\overline{p} = \sup_{x_1,x_2 \in \mathbb{X}, U_t \in \mathbb{U}^{N+1}} ||\Phi_t^+(x_1,x_2)||$ is bounded due to f and h are twice differentiable on the compact set $\mathbb{X} \times \mathbb{U}$. $\square$

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

$$d = \mathrm{col}(\xi, z) = \mathbb{T}(x) \tag{5}$$

such that the following dynamics are equivalent to (1) for any initial condition in $\mathbb{X}$ and inputs in $\mathbb{U}$,

$$\xi_{t+1} = F_1(\xi_t, z_t, u_t) \tag{6a}$$
$$z_{t+1} = F_2(z_t, u_t) \tag{6b}$$
$$y_t = g(z_t, u_t). \tag{6c}$$

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

Definition 3. The system (1) is *strongly N-detectable* if

(1) there exists a coordinate transform $\mathbb{T} : \mathbb{X} \to \mathbb{D}$ that brings the system in the form (6);
(2) the sub-system (6b)-(6c) is N-observable;
(3) the sub-system (6a) has uniformly contractive dynamics, i.e. there exists a constant $L_1 < 1$ such that

for all $\mathrm{col}(\xi_1, z) \in \mathbb{D}, \mathrm{col}(\xi_2, z) \in \mathbb{D}$ and $u \in \mathbb{U}$, the function F_1 satisfies

$$||F_1(\xi_1, z, u) - F_1(\xi_2, z, u)||' \leq L_1 ||\xi_1 - \xi_2||'. \tag{7}$$

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

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

Definition 2b. The input U_t is said to be N-*exciting* for a strongly N-detectable system (1) at time t if it is N-exciting for the corresponding sub-system (6b)-(6c) at time t.

.

The concept of N-exciting input imposes requirements that may be difficult to assess a priori. In section 3 we will study how N-excitation can be monitored online, and used in modifications to the basic MHE when this requirement is violated because the input data are not N-exciting at all times. If the input U_t is not N-exciting at certain points in time, the state estimation inversion problem (Moraal and Grizzle (1995)) will be ill-posed (the solution does not exist, is not unique, or does not depend continuously on the data) or ill-conditioned (the unique solution is unacceptably sensitive to perturbations of the data), and particular consideration is required to achieve a robust estimator. Such modifications are generally known as regularization methods, see Tikhonov and Arsenin (1977). A common method, Tikhonov and Arsenin (1977), **?**), is to augment the cost function with a penalty on deviation from a priori information and makes the estimated solution degrade gracefully when U_t is not N-exciting.[1] We utilize an explicit regularization approach that weights open loop predictions made with the underlying model (1) similar to Alessandri et al. (2008). This will have similar filtering effect as reducing the injection gain of a conventional nonlinear observer or detuning the gain of an EKF through online tuning of the process noise covariance matrix. Further regularization will be motivated later, and introduced in section 3.

A convergent estimator is pursued by the following constrained, weighted, and regularized least-squares problem of minimizing

$$J(\hat{x}_{t-N,t}, \bar{x}_{t-N}, I_t) = ||W_t(Y_t - H_t(\hat{x}_{t-N,t}))||^2 + ||M_t(\hat{x}_{t-N,t} - \bar{x}_{t-N})||^2 \tag{8a}$$

$$s.t.\ \ \hat{x}_{i+1,t} = f(\hat{x}_{i,t}, u_i), \ \ i = t-N, \ldots, t-1 \tag{8b}$$

$$\hat{x}_{t-N,t} \in \mathbb{X}, \tag{8c}$$

with M_t and W_t being time-varying weight matrices. Let $J_t^o = \min_{\hat{x}_{t-N,t}} J(\hat{x}_{t-N,t}, \bar{x}_{t-N}, I_t)$ subject to (8b)-(8c), let $\hat{x}^o_{t-N,t}$ be the associated optimal estimate, and the estimation error is defined as

$$e_{t-N} = x_{t-N} - \hat{x}^o_{t-N,t}. \tag{9}$$

It is assumed that an a priori estimator is determined as

$$\bar{x}_{t-N} = f(\hat{x}^o_{t-N-1,t-1}, u_{t-N-1}). \tag{10}$$

This formulation is a slight extension of Alessandri et al. (2008) with some additional flexibility provided by the time-varying weighting matrices W_t and M_t, which will be exploited in section 3. A condition

[1] Alternative regularization methods exist, and one implicit regularization method is to rely on the regularizing effect of an iterative approach that converges to a solution only asymptotically as $t \to \infty$ (and not converges to a solution at each individual time t), see e.g. Tautenhahn (1994). Hence, a regularizing effect is also achieved with the iterative sub-optimal variants described in Moraal and Grizzle (1995), Alessandri et al. (2008).

$W_t^T W_t > 0$ may not be sufficient for uniqueness of a solution when the input is not N-exciting. However, the condition $M_t^T M_t > 0$ is generally sufficient to guarantee that the problem has a unique solution $\hat{x}^o_{t-N,t}$. This means that the second term of (8a) can be viewed as a regularization term and the matrix M_t containing regularization parameters.

The following are assumed throughout this paper:

(A1) The set $\mathbb{U}$ is compact and convex, and the output sequence $\{y_t\}$ and the input sequence $\{u_t\}$ are bounded.

(A2) For any $\mathrm{col}(\xi_1, z_1) \in \mathbb{T}(\mathbb{X})$ and $\mathrm{col}(\xi_2, z_2) \in \mathbb{T}(\mathbb{X})$, then $\mathrm{col}(\xi_1, z_2) \in \mathbb{T}(\mathbb{X})$.

(A3) The convex and compact set $\mathbb{X}$ is controlled invariant, *i.e.* $f(x_t, u_t) \in \mathbb{X}$ for all $x_t \in \mathbb{X}$ and the control u_t for all $t \geq 0$.

(A4) The initial state $x_0 \in \mathbb{X}$, and $\bar{x}_0 \in \mathbb{X}$.

(A5) The functions f and h are twice differentiable on $\mathbb{X} \times \mathbb{U}$, and the functions F_1, F_2 and g are twice differentiable on $\mathbb{D} \times \mathbb{U}$.

(A6) $\mathbb{T}(x)$ is continuously differentiable, and bounded away from singularity for all $x \in \mathbb{X}$ such that $\mathbb{T}^{-1}(x)$ is well defined.

(A7) The system (1) is strongly N-detectable and the input U_t is N-exciting for all time $t \geq 0$.

In the stability analysis we will need to make use of the coordinate transform (5) into observable and unobservable states, although we emphasize that knowledge of this transform is not needed for the implementation of the observer. To express Y_t as a function of z_{t-N} and U_t (note that it does not depend on the unobservable states), the following algebraic mapping can be formulated similar to the mapping H:

$$Y_t = G(z_{t-N}, U_t) = G_t(z_{t-N}) = \mathrm{col}(g^{u_{t-N}}(z_{t-N}), \cdots, g^{u_t} \circ F_2^{u_{t-1}} \circ \cdots \circ F_2^{u_{t-N}}(z_{t-N})). \tag{11}$$

First, we establish lower and upper bounds on the optimal cost function J_t^o:

Lemma 2.2: *Let*

$$\Phi_t^o(z_{t-N}, \hat{z}^o_{t-N,t}) = \int_0^1 \frac{\partial}{\partial z} G((1-s)z_{t-N} + s\hat{z}^o_{t-N,t}, U_t) ds,$$
$$p_{z,t} = p_t(z_{t-N}, \hat{z}^o_{t-N,t}) = \|(W_t \Phi_t^o(z_{t-N}, \hat{z}^o_{t-N,t}))^+\|$$

Then for all $x_{t-N}, \hat{x}^o_{t-N,t} \in \mathbb{X}$

$$J_t^o \geq \left(1/p_{z,t}^2\right) \|z_{t-N} - \hat{z}^o_{t-N,t}\|^2 \tag{12}$$

Proof. Using the fact that the system (1) can be transformed using (5), there exist $d_{t-N} = \mathbb{T}(x_{t-N})$, $\hat{d}^o_{t-N,t} = \mathbb{T}(\hat{x}^o_{t-N,t})$ and $\bar{d}_{t-N} = \mathbb{T}(\bar{x}_{t-N})$ such that in the new coordinates, the system is in the form of (6a)-(6c). Note that the least squares term on the right-hand side of expression (8a) in the new coordinations can be rewritten as

$$\|W_t(Y_t - G_t(\hat{z}^o_{t-N,t}))\|^2 = \|W_t(G_t(z_{t-N}) - G_t(\hat{z}^o_{t-N,t}))\|^2.$$

From arguments similar to Lemma 2.1, it is clear that W_t can be chosen such that $p_{z,t}^2$ is uniformly bounded by any chosen positive number, and

$$\|W_t(Y_t - G(\hat{z}^o_{t-N,t}, U_t))\|^2 \geq 1/p_{z,t}^2 \|z_{t-N} - \hat{z}^o_{t-N,t}\|^2. \tag{13}$$

Taking zero as the lower bound on the second term of (8a) we get (12). □

Lemma 2.3: *Let*

$$L_2 = \max_{d\in\mathbb{D},u\in\mathbb{U}} ||\frac{\partial F_1}{\partial z}(\xi,z,u)||,\ L_3 = \max_{x\in\mathbb{X},u\in\mathbb{U}} ||\frac{\partial f}{\partial x}(x,u)||,$$

$$k_T = \max_{x\in\mathbb{X}} ||\frac{\partial \mathbb{T}}{\partial x}(x)||,\ k_{T^{-1}} = \max_{d\in\mathbb{D}} ||\frac{\partial \mathbb{T}^{-1}}{\partial d}(d)||,\ k_M = \sup_t ||M_t||$$

Then for all $x_{t-N}, \bar{x}_{t-N}, \hat{x}^o_{t-N,t} \in \mathbb{X}$

$$J^o_t \le k_M^2 L_3^2 k_{T^{-1}}^2 (||\xi_{t-N-1} - \hat{\xi}^o_{t-N-1,t-1}||^2 + ||z_{t-N-1} - \hat{z}^o_{t-N-1,t-1}||^2) \tag{14}$$

Proof. First, we remark that the Lipschitz-like constants are well defined due to (A5) and (A6) and the compactness of $\mathbb{X}$ and $\mathbb{U}$. Since (A3) and (A4) hold, x_{t-N} is a feasible solution of the MHE problem (8). From the optimality of $\hat{x}^o_{t-N,t}$, we have $J^o_t \le J(x_{t-N}, \bar{x}_{t-N}, I_t)$. It is easy to see that $\|W_t(Y_t - H(x_{t-N}, U_t))\|^2 = \|W_t(Y_t - G(z_{t-N}, U_t))\|^2 = 0$, and

$$\begin{aligned} ||M_t(x_{t-N} - \bar{x}_{t-N})||^2 &\le k_M^2 ||x_{t-N} - \bar{x}_{t-N}||^2 \\ &\le k_M^2 L_3^2 ||x_{t-N-1} - \hat{x}^o_{t-N-1,t-1}||^2 \\ &\le k_M^2 L_3^2 k_{T^{-1}}^2 (||\xi_{t-N-1} - \hat{\xi}^o_{t-N-1,t-1}||^2 + ||z_{t-N-1} - \hat{z}^o_{t-N-1,t-1}||^2). \end{aligned}$$

and the result follows. □

From the lower and upper bounds in Lemmas 2.2 and 2.3, and the inequality $\sqrt{||\xi||^2 + ||z||^2} \le ||\xi|| + ||z||$, we have for all $x_{t-N}, \hat{x}^o_{t-N,t} \in \mathbb{X}$ that

$$||z_{t-N} - \hat{z}^o_{t-N,t}|| \le q_{z,t} ||z_{t-N-1} - \hat{z}^o_{t-N-1,t-1}|| + q_{z,t} ||\xi_{t-N-1} - \hat{\xi}^o_{t-N-1,t-1}||. \tag{15}$$

with $q_{z,t} = k_M L_3 k_{T^{-1}} p_{z,t}$.

Theorem 2.4: *Let*

$$\bar{q}_z^2 = \min\left(\frac{\mu}{2(1+\mu)}, \frac{\gamma(1-L_1^2)}{2(1+\mu)}\right),\ \ \mu = \frac{1-L_1^2}{3L_1L_2},\ \ \gamma = \frac{1-L_1}{3L_2^2(1+2L_1^2)}.$$

By choosing appropriate weight matrices W_t *and* M_t *such that* $q_{z,t} \le \bar{q}_z$*, then the observer error dynamics is uniformly exponentially stable for any* $x_0, \bar{x}_0 \in \mathbb{X}$.

Proof. Since $\mathbb{X}$ is positively invariant, then $\bar{x}_{t-N} \in \mathbb{X}$, and $\bar{x}_{t-N}$ is a feasible solution. From (A2), we know $\mathrm{col}(\bar{\xi}_{t-N}, \hat{z}^o_{t-N,t})$ is also a feasible solution. Considering the cost function of the MHE problem (8), it is clear that $\mathrm{col}(\bar{\xi}_{t-N}, \hat{z}^o_{t-N,t})$ is also an optimal solution, since the first term does not depend on the unobservable states and the second term is zero, i.e. $\hat{\xi}^o_{t-N,t} = \bar{\xi}_{t-N}$. Then from (A7),

$$||\xi_{t-N} - \hat{\xi}^o_{t-N,t}|| \le L_1 ||\xi_{t-N-1} - \hat{\xi}^o_{t-N-1,t-1}|| + L_2 ||z_{t-N-1} - \hat{z}^o_{t-N-1,t-1}||. \tag{16}$$

Let $s_{1,t} = \|z_{t-N} - \hat{z}^o_{t-N,t}\|$, and $s_{2,t} = \|\xi_{t-N} - \hat{\xi}^o_{t-N,t}\|$. Then combining (15) and (16) gives

$$\begin{pmatrix} s_{1,t} \\ s_{2,t} \end{pmatrix} \leq \begin{pmatrix} q_{z,t}, & q_{z,t} \\ L_2, & L_1 \end{pmatrix} \begin{pmatrix} s_{1,t-1} \\ s_{2,t-1} \end{pmatrix}. \tag{17}$$

Since $s_{i,t} \geq 0$, it follows that $s_{i,t} \leq \bar{s}_{i,t},\ \ i = 1,2$, where we define the 2nd order linear time-varying system

$$\begin{pmatrix} \bar{s}_{1,t} \\ \bar{s}_{2,t} \end{pmatrix} = \begin{pmatrix} q_{z,t}, & q_{z,t} \\ L_2, & L_1 \end{pmatrix} \begin{pmatrix} \bar{s}_{1,t-1} \\ \bar{s}_{2,t-1} \end{pmatrix}, \ \begin{pmatrix} \bar{s}_{1,0} \\ \bar{s}_{2,0} \end{pmatrix} = \begin{pmatrix} s_{1,0} \\ s_{2,0} \end{pmatrix}.$$

Consider a Lyapunov function candidate $V(\bar{s}_1, \bar{s}_2) = \bar{s}_1^2 + \gamma \bar{s}_2^2$ with $\gamma > 0$. It follows that

$$\begin{aligned} & V(\bar{s}_{1,t}, \bar{s}_{2,t}) - V(\bar{s}_{1,t-1}, \bar{s}_{2,t-1}) \\ & \quad = -(1 - q_{z,t}^2 - \gamma L_2^2)\bar{s}_{1,t-1}^2 - (\gamma - q_{z,t}^2 - \gamma L_1^2)\bar{s}_{2,t-1}^2 + 2(q_{z,t}^2 + \gamma L_1 L_2)\bar{s}_{1,t-1}\bar{s}_{2,t-1}. \end{aligned}$$

According to Young's inequality, for any $\mu > 0$,

$$\begin{aligned} & V(\bar{s}_{1,t}, \bar{s}_{2,t}) - V(\bar{s}_{1,t-1}, \bar{s}_{2,t-1}) \\ & \quad \leq -(1 - q_{z,t}^2 - \gamma L_2^2)\bar{s}_{1,t-1}^2 - (\gamma - q_{z,t}^2 - \gamma L_1^2)\bar{s}_{2,t-1}^2 + (q_{z,t}^2 + \gamma L_1 L_2)/\mu \bar{s}_{1,t-1}^2 + (q_{z,t}^2 + \gamma L_1 L_2)\mu \bar{s}_{2,t-1}^2 \\ & \quad \leq -\delta_1 \bar{s}_{1,t-1}^2 - \delta_2 \bar{s}_{2,t-1}^2. \end{aligned}$$

where

$$\delta_1 = 1 - (1 + 1/\mu) q_{z,t}^2 - \gamma \left(L_2^2 + L_1 L_2/\mu \right) \tag{18}$$

$$\delta_2 = \gamma(1 - L_1^2) - (1 + \mu) q_{z,t}^2 - \gamma L_1 L_2 \mu \tag{19}$$

First, choose μ such that $L_1 L_2 \mu = \frac{1}{3}(1 - L_1^2)$. Then the first term of δ_2 dominates its third term by a factor 3, and

$$\mu = \frac{1 - L_1^2}{3 L_1 L_2} > 0. \tag{20}$$

Second, choose γ such that $\gamma(L_1 L_2/\mu + L_2^2) = \frac{1}{3}$, which leads to the first term of δ_1 dominating its third term by a factor 3, and

$$\gamma = \frac{1 - L_1^2}{3 L_2^2 (1 + 2 L_1^2)} > 0. \tag{21}$$

Third, since $q_{z,t}$ is chosen such that the first terms of both δ_1 and δ_2 dominate their second terms by a factor 2, respectively, and we have

$$(1+1/\mu)q_{z,t}^2 \leq \frac{1}{2} \;\Rightarrow q_{z,t}^2 \leq \frac{\mu}{2(1+\mu)},$$
$$(1+\mu)q_{z,t}^2/\gamma \leq \frac{1}{2}(1-L_1^2) \;\Rightarrow q_{z,t}^2 \leq \frac{\gamma(1-L_1^2)}{2(1+\mu)},$$

such that $\delta_1 > 0$ and $\delta_2 > 0$. There always exists a matrix M_t with some sufficiently small k_M and a matrix W_t for some sufficiently small $p_{z,t}$ such that $q_{z,t} \leq \bar{q}_z$ such that $\delta_1 > 0$ and $\delta_2 > 0$, and the 2nd order LTV system is uniformly exponentially stable for the given initial conditions. Since $s_{i,t} \leq \bar{s}_{i,t}$ and (A6) holds, the error dynamics is also uniformly exponentially stable for any $x_0, \bar{x}_0 \in \mathbb{X}$. $\square$

Assumption (A2) is used in the proof to ensure that a feasible solution $\bar{x}_{t-N}$ remains feasible in the transformed coordinates when the observable states are replaced by their optimal values. This assumption is trivially satisfied for any N-observable system. For systems that are not N-observable, but N-detectable, it will still hold trivially in many cases as illustrated in Example 1 later. Like many other assumptions in this paper, such as (A7), it will not be trivial to verify unless $\mathbb{T}$ is known. However, Theorem 2.4 remains of value in such cases since it provides a qualitative understanding of the method. Hence, the theory provides a guideline, rather than replacement, for practical tuning as illustrated in the examples.

If the data are not N-exciting, the second term of the observer cost function dominates and the observer degenerates to an open loop observer for the state combinations that are not excited, provided M_t has full column rank in the sub-space corresponding to the linear combination of states not being excited. With $M_t^T M_t > 0$ this is trivially satisfied. This may be a satisfactory solution if the system has open loop asymptotically stable dynamics within the region of operation, since the observer may still converge and give accurate estimates. In practise, the accuracy will then depend entirely on the accuracy of the model. If the system is not open loop asymptotically stable, and in particular if there are significant model errors, this approach may not be satisfactory since errors will be allowed to accumulate without the presence of feedback from measurements. This will be the case in a mixed parameter and state estimation problem with the state space $x = \mathrm{col}(\chi, \theta)$ corresponds to the system state χ and the unknown parameters θ and the augmented dynamics

$$\chi_{t+1} = f(\chi_t, \theta_t, u_t) \tag{22}$$
$$\theta_{t+1} = \theta_t \tag{23}$$
$$y_t = h(\chi_t, \theta_t) \tag{24}$$

Regardless of the system dynamics f, the augmented parameter dynamics $\theta_{t+1} = \theta_t$ are not asymptotically stable and estimates may drift off due to integrated errors (see Example 1 later). In the next section we introduce further methods for weighting and regularization that degrade gracefully when data are not N-exciting, which are particularly useful when the system is not asymptotically stable, as in the case with mixed state and parameter estimation, and there are model errors.

3 Adaptive weighting and regularization without persistence of excitation

In order to implement excitation-sensitive regularization, it is essential to be able to monitor if the data are N-exciting or not. For N-observable systems, the condition $p_{z,t} \leq \bar{q}_z/(k_M L_3 k_{T-1})$ will depend on the existence of a (not too small) $\varepsilon > 0$ such that

$$\Phi_t^T(x_{t-N},\hat{x}^o_{t-N,t})\Phi_t(x_{t-N},\hat{x}^o_{t-N,t}) \geq \varepsilon I > 0 \tag{25}$$

for all $t \geq 0$, where

$$\Phi_t(x_1,x_2) = \int_0^1 \frac{\partial}{\partial x} H((1-s)x_1 + sx_2, U_t)ds \tag{26}$$

This condition comes from the requirement of U_t being N-exciting at all t and is similar to a PE condition. Unfortunately, since $\Phi_t(x_{t-N},\hat{x}^o_{t-N,t})$ depends on the unknown x_{t-N} we cannot compute $p_t(x_{t-N},\hat{x}^o_{t-N,t}) = \|(W_t\Phi_t(x_{t-N},\hat{x}^o_{t-N,t}))^+\|$ exactly at any point in time to monitor if U_t is N-exciting. Instead, we have to rely on some approximation or estimate of $p_t(\cdot)$. If it is assumed that $||e_{t-N}||$ is small, then

$$\Phi_t(x_{t-N},\hat{x}^o_{t-N,t}) \approx \Phi_t(\hat{x}^o_{t-N,t},\hat{x}^o_{t-N,t}) = \frac{\partial H}{\partial x}(\hat{x}^o_{t-N,t},U_t)$$

and we can use $\hat{p}_t(\hat{x}^o_{t-N,t}) = \|(W_t\frac{\partial H}{\partial x}(\hat{x}^o_{t-N,t},U_t))^+\|$ to approximate $p_t(x_{t-N},\hat{x}^o_{t-N,t})$. Consider a singular value decomposition (SVD), Golub and van Loan (1983)

$$\frac{\partial H}{\partial x}(\hat{x}^o_{t-N,t},U_t) = U_tS_tV_t^T. \tag{27}$$

Any singular value (diagonal element of the matrix S_t) that is zero or close to zero indicates that a state component is unobservable or the input is not N-exciting. Moreover, the corresponding row of the V_t matrix will indicate which components cannot be estimated. The Jacobian has the structural property that its rank will be no larger than $\dim(z) = n_z \leq n_x$, due to certain components being unobservable. In addition, its rank can be reduced by data being not N-exciting as discussed in Lemma 2.1. The N-excitation of data may therefore be monitored through the robust computation of the rank of the Jacobian matrix using the SVD, Golub and van Loan (1983).

We know that convergence depends on W_t being chosen such that $p_{z,t}$ is bounded by a sufficiently small number. To pursue this objective, one may choose W_t such that, whenever possible,

$$\|(W_t\Phi_t(\hat{x}^o_{t-N,t},\hat{x}^o_{t-N,t}))^+\| = \alpha, \tag{28}$$

where $\alpha > 0$ is a sufficiently small scalar. In order to give zero weight on data for components that are either unobservable or unexcited, we modify this ideal design equation into the more practical and realistic design objective

$$\|(W_tU_tS_{t,\delta}V_t^T)^+\| = \begin{cases} \alpha, \text{ if } ||S_t|| \geq \delta \\ 0, \text{ otherwise} \end{cases} \tag{29}$$

where the thresholded pseudo-inverse $S^+_{t,\delta} = \mathrm{diag}(1/\sigma_{t,1},\ldots,1/\sigma_{t,\ell},0,\ldots,0)$ where $\sigma_1,\ldots,\sigma_\ell$ are the singular values larger than some $\delta > 0$ and the zeros correspond to small singular values whose inverse is set to zero, Golub and van Loan (1983). This leads to

$$W_t = (1/\alpha)V_tS^+_{t,\delta}U_t^T \tag{30}$$

satisfying

$$||(W_t\Phi_t(\hat{x}^o_{t-N,t},\hat{x}^o_{t-N,t}))^+|| \leq \alpha. \tag{31}$$

Here M_t is chosen as

$$M_t = \beta I_{n_x}, \tag{32}$$

where $\beta \geq 0$ is a scalar.The following result shows that the W_t defined in (31) satisfies the conditions of Theorem 2.4 locally.

Theorem 3.1: *If W_t is chosen according to (30) with δ being sufficiently small and $0 < \alpha < \bar{q}_z/(L_3 k_{T^{-1}} k_T \beta)$, then W_t is bounded and the observer error dynamics is locally uniformly exponentially stable.*

Proof. Boundedness of W_t follows directly from $\alpha, \delta > 0$. Since δ is sufficiently small and the data are N-exciting, we assume without loss of generality that the matrix $W_t\Phi^o_t(z_{t-N},\hat{z}^o_{t-N,t})$ has full rank. Using similar arguments as Lemma 2.1, it is easy to show that (13) in the proof of Lemma 2.2 is still valid.

$$\begin{aligned}Y - H(\hat{x}^o_{t-N,t},U_t) &= \Phi_t(x_{t-N},\hat{x}^o_{t-N,t})(x_{t-N}-\hat{x}^o_{t-N,t}),\\ Y - G(\hat{z}^o_{t-N,t},U_t) &= \Phi^o_t(z_{t-N},\hat{z}^o_{t-N,t})(z_{t-N}-\hat{z}^o_{t-N,t}).\end{aligned}$$

To simplify the notation, let $\Phi_{x,t} = \Phi_t(x_{t-N},\hat{x}^o_{t-N,t})$ and $\Phi_{z,t} = \Phi^o_t(z_{t-N},\hat{z}^o_{t-N,t})$. Since $Y - H(\hat{x}^o_{t-N,t},U_t) = Y - G(\hat{z}^o_{t-N,t},U_t)$,

$$\Phi_{z,t}(z_{t-N}-\hat{z}^o_{t-N,t}) = \Phi_{x,t}(x_{t-N}-\hat{x}^o_{t-N,t}).$$

It is known that

$$d_{t-N}-\hat{d}^o_{t-N,t} = \Gamma_t(x_{t-N}-\hat{x}^o_{t-N,t}),$$

where $\Gamma_t = \int_0^1 \frac{\partial}{\partial x}\mathbb{T}((1-s)x_{t-N}+s\hat{x}^o_{t-N,t})ds$. Together with $z = \eta d$, where $\eta = [\mathbf{0}_{n_z\times(n_x-n_z)}, I_{n_z\times n_z}]$, we have

$$\Phi_{z,t}\eta(d_{t-N}-\hat{d}^o_{t-N,t}) = \Phi_{x,t}\Gamma_t^{-1}(d_{t-N}-\hat{d}^o_{t-N,t}).$$

It follows that

$$\begin{aligned}&W_t\Phi_{z,t}\eta = W_t\Phi_{x,t}\Gamma_t^{-1}\\ \Rightarrow&(W_t\Phi_{z,t})^+ = \eta\Gamma_t(W_t\Phi_{x,t})^+\\ \Rightarrow&||(W_t\Phi_{z,t})^+|| \leq ||\eta||\cdot||\Gamma_t||\cdot||(W_t\Phi_{x,t})^+||.\end{aligned} \tag{33}$$

It is known that $||\eta|| = 1$ and $||\Gamma_t|| \leq k_T$. When $||e_t||$ belongs to a neighborhood of the origin, from (31), we know that $||(W_t\Phi_{x,t})^+|| \leq \alpha$. Then

$$p_{z,t} = ||(W_t\Phi_{z,t})^+|| \leq k_T\alpha.$$

From Theorem 2.4, to obtain conditions on the Lyapunov function, $q_{z,t} = k_{T-1}\beta L_3 p_{z,t} \leq \bar{q_z}$. Therefore, the convergence condition is implied by $\alpha < \bar{q_z}/(L_3 k_{T-1} k_T \beta)$. Note that (33) and the following arguments only holds for $||e_t||$ in a neighborhood of the origin, therefore only local exponential convergence results. □

The tuning parameters with this adaptive choice of W_t and M_t are the non-negative scalars α, δ and β. It is worthwhile to notice that since they are scalars, a successful tuning of the observer will depend on appropriate scaling of the variables and model equations. Basic scaling should be made by normalization such that all variables have approximately the same range and magnitude. Next, individual performance requirements for the individual state estimates can be achieved by increasing or decreasing the individual scaling parameters. For example, this is necessary to tune the degree of filtering done by the observer in order to match the individual sensor noise levels and uncertainty levels in the model equations.

When the data are not considered N-exciting at some time instant, then δ should be tuned such that the corresponding singular values of $\Phi_t(\cdot)$ will be less than δ such that W_t defined by (30) will not have full rank. The means that the error in the corresponding state combinations will not be penalized by the first term in the criterion, and due to the second term the estimates will be propagated by the open loop model dynamics. In case of augmented parameter states $\theta_{t+1} = \theta_t$ this means that they are essentially frozen at their value from the previous sample. It is clear that appropriate scaling is essential, as mentioned above, otherwise the singular values cannot be compared and it might be infeasible to find an appropriate setting for the threshold δ that is suitable for all situations.

Both $\beta \geq 0$ and $\delta \geq 0$ could be considered as regularization parameters that must be chosen carefully in order to tune the practical performance of the observer. In the ideal case with a perfect model, no noise, no disturbances and N-exciting data at all sampling instant one could choose $\delta = \beta = 0$. As a practical tuning guideline we propose to first choose $\beta > 0$ in order to achieve acceptable filtering and performance with typical noise and disturbance levels for typical cases when the data are N-exciting. Second, $\delta > 0$ is chosen in order to achieve acceptable performance also in operating conditions when the data are not N-exciting. In both these steps, one should also regard the scaling of the variables and equations as individual tuning parameters for fine-tuning, in addition to the mentioned scalar tuning parameters.

4 Example

4.1 *Example 1 - mixed state and parameter estimation*

Consider the following nonlinear system

$$\dot{x}_1 = -2x_1 + x_2 \tag{34a}$$

$$\dot{x}_2 = -x_2 + x_3(u - w) \tag{34b}$$

$$\dot{x}_3 = 0 \tag{34c}$$

$$y = x_2 + v. \tag{34d}$$

One may think of x_3 as a parameter representing an unknown gain on the input, where the third state equation is an augmentation for the purpose of estimating this parameter. It is clear that x_1 is not observable, but corresponds to an asymptotically stable sub-system. It is also clear that the observability of x_3 will depend on the excitation u, while x_2 is generally observable.

The same observability and detectability properties hold for the discretized system with sampling interval $t_f = 0.1$. When $u = 0$ for all time, the rank of $\frac{\partial H}{\partial x}(\hat{x}^o_{t-N,t}, U_t)$ is 1. When u is white noise, the rank of $\frac{\partial H}{\partial x}(\hat{x}^o_{t-N,t}, U_t)$ is 2 almost always.

In this simulation example we choose $N = 2$ such that the moving window has length $N+1 = 3$. The base case is defined as follows. We use the adaptive weighting law (30) with $\alpha = 1$, $\delta = 0.1$. Measurement noise, with independent uniformly distributed $v \in [-0.5, 0.5]$, is added to the base case. The input is chosen with periods without informative data as follows: During $0 \leq t < 30t_f$, $u = 0$. During $30t_f \leq t < 60t_f$, u is discrete-time white noise. During $60t_f \leq t \leq 120t_f$, $u = 0$. In the simulation, true system has an input disturbance with $w = 0.15$, and the model used in the MHE observer (8b) with no explicit knowledge of the input disturbance. In the following figures, true states are shown in solid line; estimated states of proposed work are shown in dash-dot line; estimated states using the alternative setting with fixed W_t are shown in dash line. The following initial conditions are used: $x_0 = [4, -7, 2], \bar{x_0} = [3, -5.9, -1]$.

- Case 1: Choose $\beta = 1$ for the proposed observer. For the fixed weight alternative setting, choose $W_t = 4I$ and $\beta = 1$. The simulation result is shown in Figure 1.
- Case 2: Choose $\beta = 0$ for the proposed observer. For the fixed weight alternative setting, choose $W_t = I$ and $\beta = 0$. The simulation result is shown in Figure 2.

The example shows that the adaptive weighting with the thresholded singular value inversion effectively freezes the unexcited parameter estimate and thereby avoids the parameter estimate drift that otherwise may result due to unmatched model error (input disturbance) when there are no excitations. Additional regularization is achieved by $\beta > 0$ since otherwise the parameter estimation will be mainly dominated by noise, as shown by case 2.

In the following, several values of α, δ and β are chosen in order to illustrate how the performance depends on these choices. Let us consider the performance indexes given by the root mean square error: $RMSE = (\sum_{t=0}^{total} \frac{||e_t||^2}{total})^{1/2}$, where $||e_t||$ is the norm of the estimation error at time t, $total$ is the length of the simulation run. We choose $total = 120$ and Table 1 presents the performances of choosing different parameters. We observe that the chosen tuning in Figures 1 and 2 appears to be a good one, while at the same time performance does not degrade very strongly when the tuning is changed. This indicates that the method is somewhat robust to the choice of tuning parameters.

α	1	1	1	1	1	1	10	10	10	0.1
δ	0.1	0.1	0.1	1	10	0.01	0.1	0.1	1	0.1
β	1	10	0.1	1	1	1	1	10	1	1
RMSE	0.1162	0.3082	0.1196	0.1139	0.3221	0.1216	0.2823	0.3538	0.2922	0.1208

Table 1. The performances depending on the tuning parameters.

4.2 *Example 2 - Wheel slip and tyre-road friction estimation*

Consider the longitudinal dynamics corresponding to one wheel and 1/4 of a vehicle mass (quarter-car model). The wheel slip dynamics are important in the control algorithm of an anti-lock brake systems

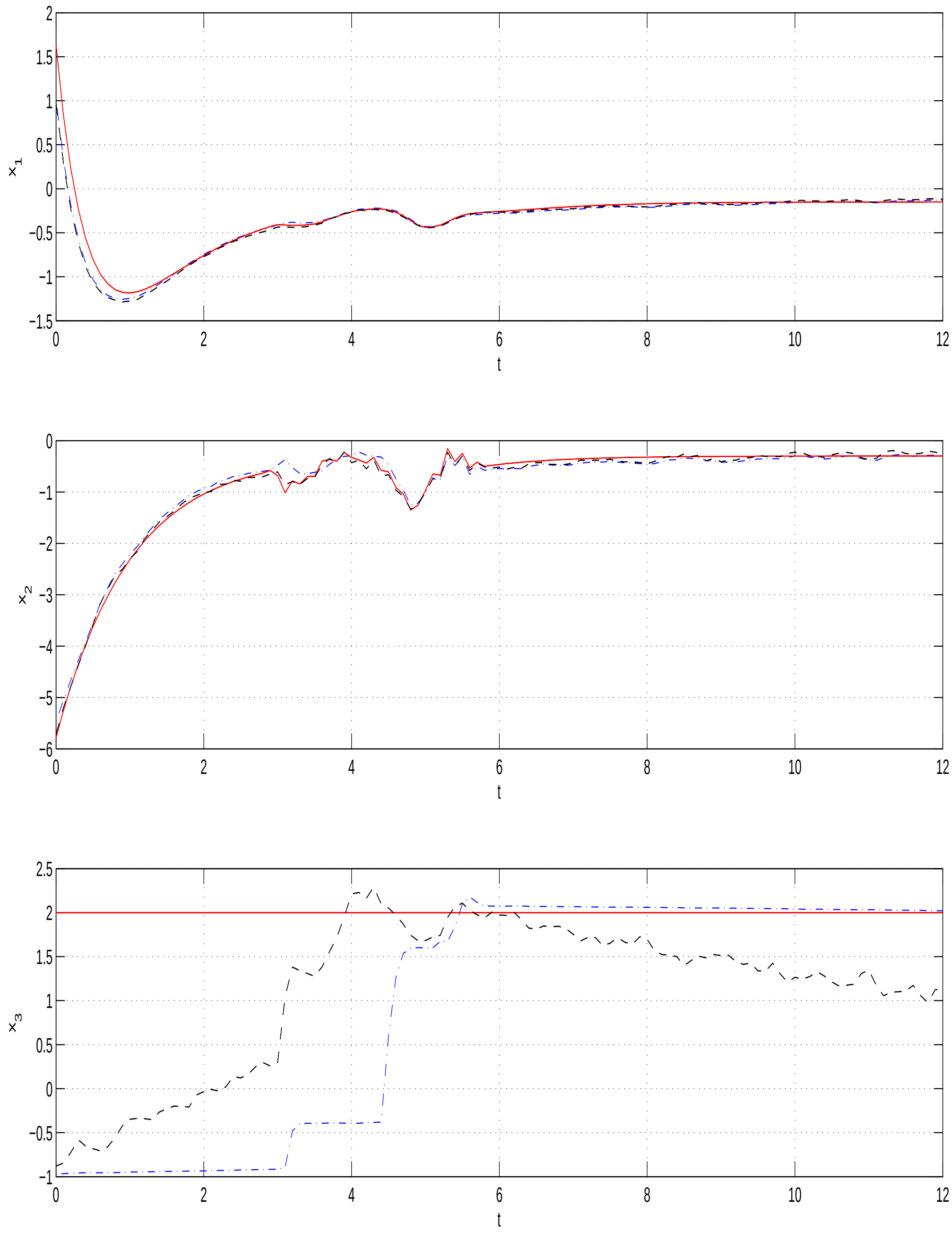

Figure 1. Simulation results of case 1.

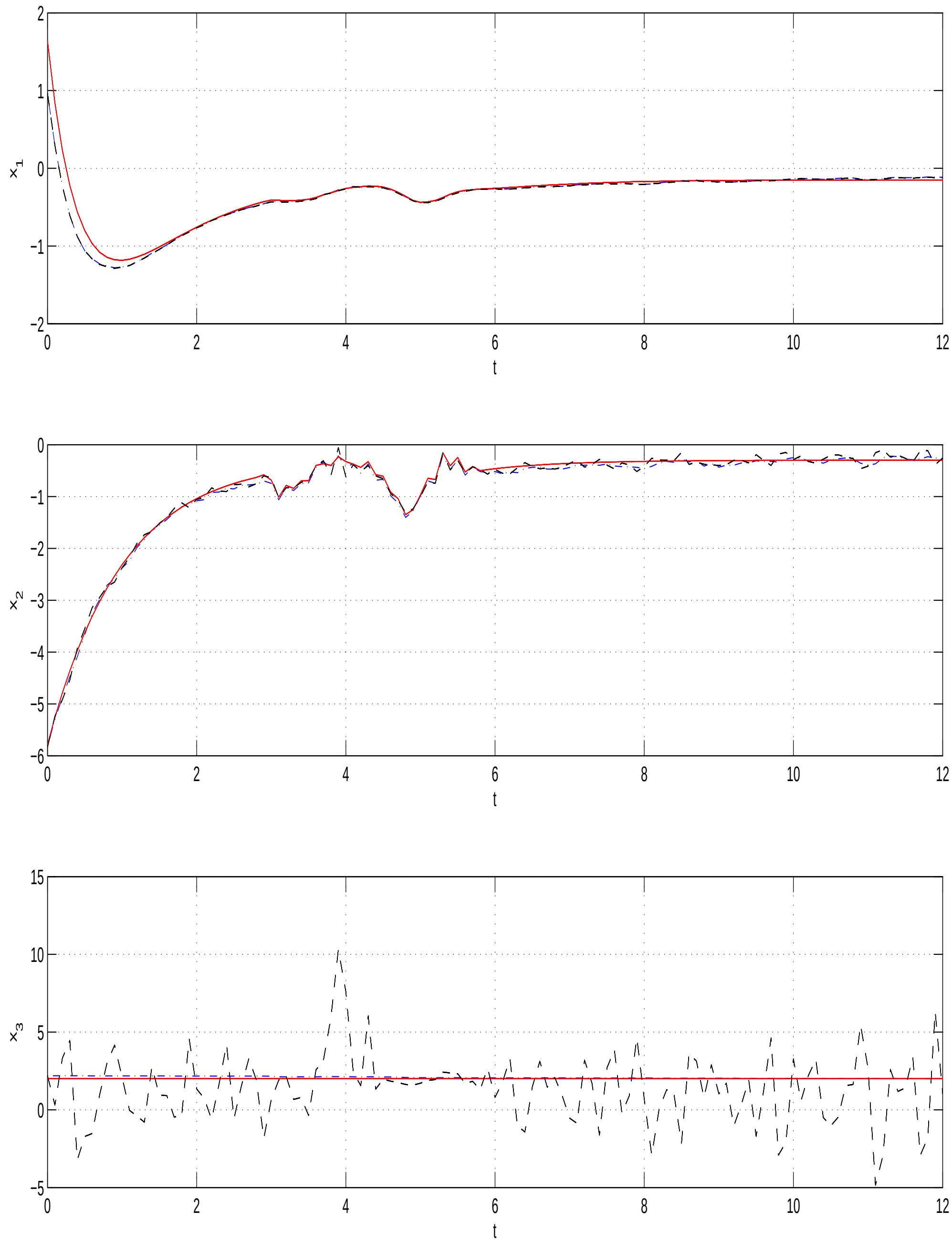

Figure 2. Simulation results of case 2.

(ABS), Burckhardt (1993), Johansen et al. (2003)

$$\dot{v} = -\frac{1}{m} F_z \mu_1(\lambda), \tag{35a}$$

$$\dot{\lambda} = -\frac{1}{v}\left(\frac{1}{m}(1-\lambda) + \frac{r^2}{J}\right) F_z \mu_1(\lambda) + \frac{1}{v}\frac{r}{J} T_b, \tag{35b}$$

$$y = \frac{v(1-\lambda)}{r}, \tag{35c}$$

where v is the longitudinal speed, λ is longitudinal tyre slip, T_b is torque acting on wheel axis, $F_z = mg$ is the vertical force, μ_1 is the friction coefficient, and y is the angular wheel speed measurement. This application requires a combined state and parameter estimator since only the angular wheel speed y is measured, such that both λ and v needs to be estimated togehter with parameters of the tire-road friction model that defines the friction coefficient μ_1. In the example the following constant parameters are known $m = 325$, $r = 0.345$, $J = 1$ and $g = 9.81$, while the friction coefficient $\mu_1(\lambda)$ is a nonlinear function of the longitudinal slip λ with

$$\mu_1(\lambda) = \theta \sin(C(\arctan B\lambda - E(B\lambda - \arctan(B\lambda)))). \tag{36}$$

and the parameters B, C, E and θ characterize the tire and the road surface. Typical values of parameters B, C, E and θ are given by Matuško et al. (2003)

$$\text{Dry asphalt:}\;\; B = 10.38,\; C = 1.65,\; E = 0.65663,\; \theta = 1 \tag{37}$$

$$\text{Snow:}\;\; B = 14.395,\; C = 0.9,\; E = -6.439,\; \theta = 0.3 \tag{38}$$

In this example, it is assumed that the parameters B, C, E, θ are unknown. With this parameterization one has to expect that the model will be over-parameterized such that the persistence of excitation condition (and uniform observability) will not hold. This challenging parameterization is chosen in order to illustrate the power of the proposed method, and in particular that the algorithm will accurately detect the excitation level of the data at any time and adapt the weights accordingly when using W_t defined by (30). Therefore, the proposed MHE algorithm is applied to the combined state and parameter estimation problem. Considering the parameters θ, E, C, B as augmented states, the states B, C, E, θ are added,

$$\dot{B} = 0,\; \dot{C} = 0,\; \dot{E} = 0,\; \dot{\theta} = 0. \tag{39}$$

The constraints on the states using the the MHE optimization are given as

$$\begin{aligned} 1 \le & v(t) \le 30, \\ 0 \le & \lambda(t) \le 1, \\ 0 \le & \theta(t) \le 1, \\ 9 \le & B(t) \le 15.5, \\ 0 \le & C(t) \le 3, \\ -7.5 \le & E(t) \le 2. \end{aligned}$$

and the system is discretized using the standard Euler method. We remark that the lower bound on $v(t)$ is conventional, since the ABS application will handle low speed as an exception where controllability is lost due to the singularity at $v = 0$ Johansen et al. (2003). In the example we choose the initial conditions $v(0) = 20, \lambda(0) = 0.01$, and the true values θ, B, C and E are given in (37) and (38) according to the different scenarios. In the simulation, choose the initial a priori estimates $\bar{v}(0) = 19, \bar{\lambda}(0) = 0, \bar{\theta}(0) = 0.6$, $\bar{B}(0) = 12$, $\bar{C}(0) = 1.3$ and $\bar{E}(0) = 0$. The horizon is chosen as $N = 10$. The sampling interval $t_f = 0.01$ s, and Gaussian white noise with variance 0.2 rad/s is applied to the measurements. Choose $\alpha = 0.01$ and $\beta = 1$. W_t is chosen with $\delta = 0.8$.

- For dry asphalt, the simulation result is shown in Figure 3.
- For snow, the simulation result is shown in Figure 4.

In the figures, true states are shown in solid lines and estimated states are shown in dash lines. It is interesting to observe that the estimation is robust and that the SVD thresholding effectively prevents the estimates of (B, C, E) from drifting and becoming highly incorrect when there is not much excitation or they are poorly observable. There are slightly more excitations in the dry asphalt case (stronger braking and higher wheel slips) and the adaptive weighting makes more attempts to estimate the parameters C and E in this case, compared to the show scenario. The parameter B is in both cases not excited, while good estimates of the most important variables λ and θ are achieved in both scenarios.

5 Discussion and conclusions

Theoretical and practical properties of a regularized nonlinear moving horizon observer were demonstrated in this paper. Although no convergence problems due to local minima were encountered in the simulation example in this paper, it is important to have in mind that the method will rely on a sufficiently accurate guess of the initial a priori estimate in cases when sub-optimal local minima exist.

The main feature of the proposed method is systematic handling of nonlinear systems that are neither uniformly observable, nor persistently excited, and may not be asymptotically stable. This is a typical situation with mixed parameter and state estimation with an augmented state space model. With the exception of the preliminary results in Moraal and Grizzle (1995), this is to the best of the authors knowledge, an important issue not studied in depth in any other nonlinear moving horizon observer. The examples show that the method can be successfully tuned and applied in challenging cases when the uniform observability and persistence of excitation conditions are not fulfilled, even with a highly over-parameterized model, without the need for careful a priori analysis of observability and persistence of excitation conditions. By proper scaling and tuning, the algorithm can automatically adapt to the level of excitation. An industrial example is found in Paasche et al. (2011).

Acknowledgements

This research was supported by the Research Council of Norway under the Strategic University Program on Computational Methods in Nonlinear Motion Control.

References

Abraham, R., Marsden, J.E., and Ratiu, T., *Manifolds, Tensor Analysis, and Applications*, Springer-Verlag New York (1983).

Alamir, M. (1999), “Optimization based non-linear observers revisited,” *Int. J. Control*, 72, 1204–1217.

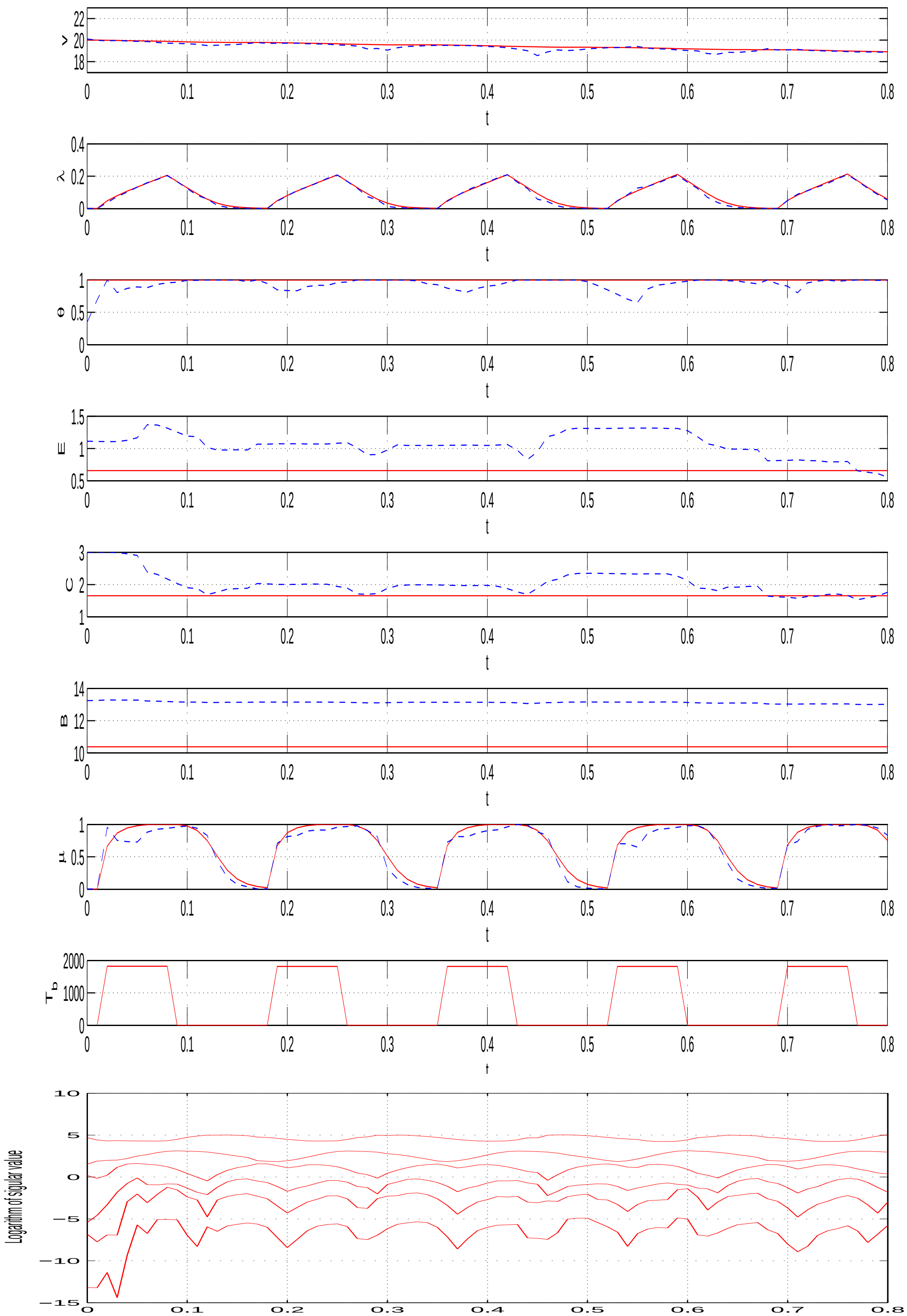

Figure 3. Example 2: Simulation results with dry asphalt road conditions.

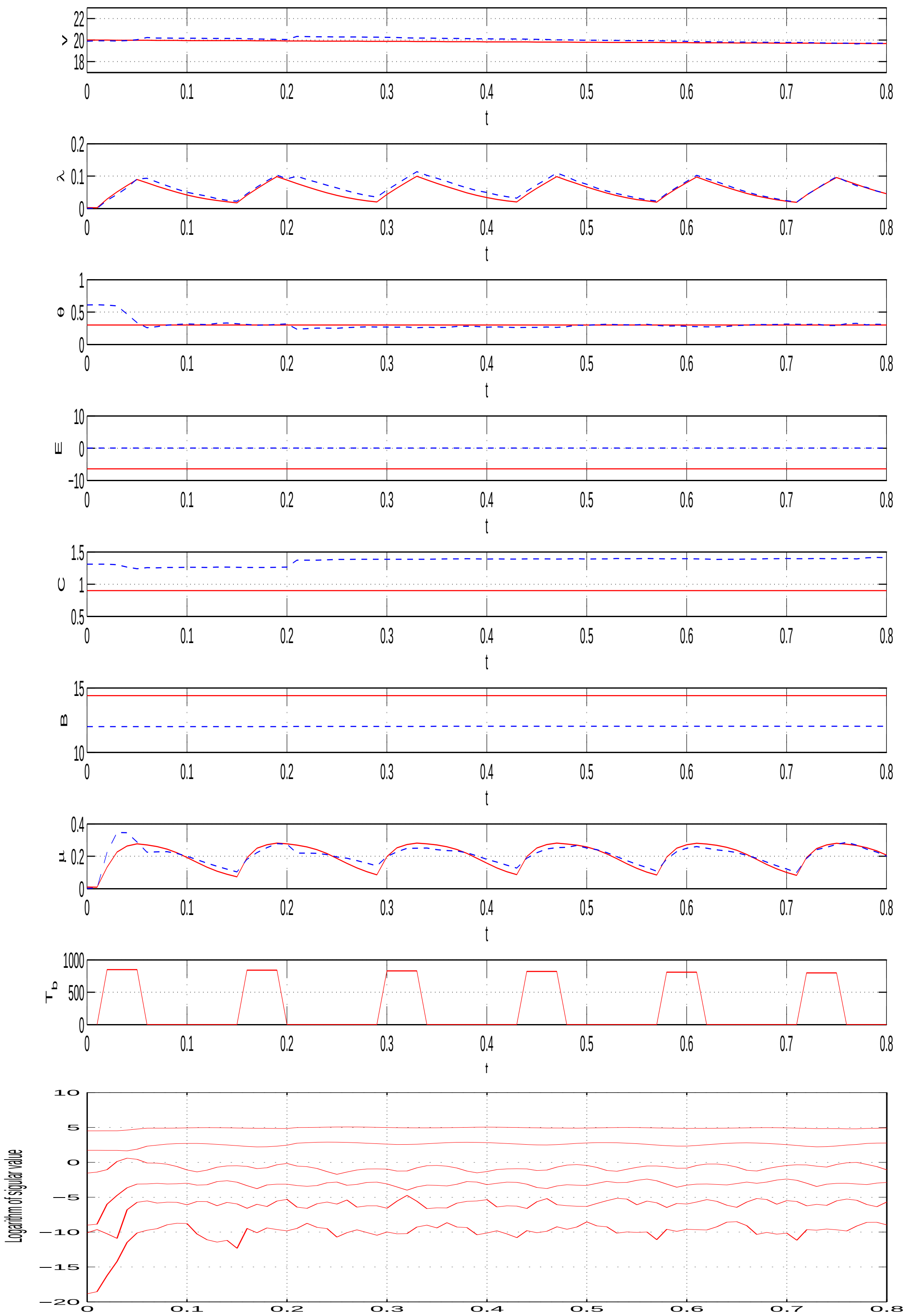

Figure 4. Example 2: Simulation results with snow road conditions.

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

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

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

Burckhardt, M. (1993), "Fahrwerktechnik: Radschlupf-Regelsysteme," *Vogel Verlag*.

Campi, M.C. (1994), "Exponentially Weighted Least/Squares Identification of Time-Varying Systems with White Disturbances," *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 42.

Cao, L.Y., and Schwartz, H.M. (1999), "A Directional Forgetting Algorithm Based on the Decomposition of the Information Matrix," in *Proceedings of the 7th Mediterranean Conference on Control and Automation, Haifa, Israel*.

Fiacco, A.V., *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press (1983).

Fortescue, T.R., Kershenbaum, L., and Ydstie, B. (1981), "Implementation of Self-tuning Regulators with Variable Forgetting Factors," *Automatica*, 17, 831835.

Gelb, A., *Applied Optimal Estimation*, 17 ed., MIT Press (2002).

Glad, S.T. (1983), "Observability and nonlinear dead beat observers," in *IEEE Conf. Decision and Control, San Antonio*, pp. 800–802.

Golub, G.H., and van Loan, C.F., *Matrix computations*, Oxford University Press (1983).

Grossman, W.D. (1999), "Enhancing Noise Robustness in Discrete-time nonlinear observers," in *Proc. American Control Conference, San Diego*, pp. 3489–3494.

Haseltine, E.L., and Rawlings, J.B. (2005), "Critical evaluation of Extended Kalman Filtering and Moving-Horizon Estimation," *Ind. Eng. Chem. Res*, 44, 2451–2460.

Hovd, M., and Bitmead, R.R. (2007), "Directional Leakage and Parameter Drift," *Int. J. Adaptive. Control and Signal Processing*, 20, 27–39.

Johansen, T.A. (1996), "Identification of Non-linear Systems using Empirical Data and Prior Knowledge - An Optimization Approach," *Automatica*, 32, 337–356.

Johansen, T.A., Petersen, I., Kalkkuhl, J., and Ludemann, J. (2003), "Gain-scheduled wheel slip control in automotive brake systems," *IEEE Trans. Control Systems Technology*, 11, 799–811.

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

Kulhavy, R. (1987), "Restricted Exponential Forgetting in Real-time Identication," *Automatica*, 23, 589–600.

Matuško, J., Petrović, I., and Perić, N. (2003), "Application of Extended Kalman Filter for Road Condition Estimation," *Automatika*, 44, 59–65.

Moraal, P.E., and Grizzle, J.W. (1995), "Asymptotic observers for detectable and poorly observable systems," in *IEEE Conf. Decision and Control, New Orleans*, pp. 109–114.

Moraal, P.E., and Grizzle, J.W. (1995), "Observer Design for Nonlinear Systems with Discrete-time Measurement," *IEEE Transactions Automatic Control*, 40, 395–404.

Paasche, M., Johansen, T.A., and Imsland, L.S. (2011), "Estimation of Bottomhole Pressure during Oil Well Drilling Using Regularized Adaptive Nonlinear Moving Horizon Observer," in *IFAC World Congress, Milano, Italy*.

Panteley, E., Loria, A., and Teel, A. (2001), "Relaxed persistency of excitation for uniform asymptotic stability," *IEEE Trans. Automatic Control*, 46, 1874–1886.

Raff, T., Ebenbauer, C., Findeisen, R., and Allgöwer, F. (2005), "Remarks on Moving Horizon State Estimation with Guaranteed Convergence," in *Control and Obsever Design for Nonlinear Finite and Infinite Dimensional Systems* eds. T. Meurer, K. Graichen and E.D. Gilles, Springer-Verlag, Berlin, pp. 67–80.

Rao, C.V., Rawlings, J.B., and Mayne, D.Q. (2003), “Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximation,” *IEEE Transactions Automatic Control*, 48, 246–258.

Reif, K., Sonnemann, F., and Unbehauen, R. (1998), “EKF-Based Nonlinear Observer with a Prescribed Degree of Stability,” *Automatica*, 34, 1119–1123.

Reif, K., and Unbehauen, R. (1999), “The Extended Kalman Filter as an Exponential Observer for Non-linear Systems,” *IEEE Transactions on Signal Processing*, 47, 2324–2328.

Salgado, M., Goodwin, G., and Middleton, R. (1988), “Modified Least Squares Algorithm Incorporating Exponential Resetting and Forgetting,” *Int. J. Control*, 47, 477 – 491.

Sedoglavic, A. (2002), “A probabilistic algorithm to test local algebraic observability in polynomial time,” *J. Symbolic Computation*, 33, 735755.

Sui, D., and Johansen, T.A. (2010), “Regularized Nonlinear Moving Horizon Observer for Detectable Systems,” in *IFAC NOLCOS, Bologna, Italy*.

Tautenhahn, U. (1994), “On the asymptotic regularization of nonlinear ill-posed problems,” *Inverse Problems*, 10, 1405–1418.

Tikhonov, A.N., and Arsenin, V.Y., *Solutions of Ill-posed Problems*, Wiley (1977).

Zimmer, G. (1994), “State observation by on-line minimization,” *Int. J. Control*, 60, 595–606.