Convergence guarantees for moving horizon estimation based on the real-time iteration scheme

A. Wynn¹, M. Vukov², M. Diehl².

Abstract—In this note, conditions are proven under which a real-time implementable moving horizon estimation (MHE) scheme is locally convergent. Specifically, the real-time iteration scheme of [17] is studied in which a single Gauss-Newton iteration is applied to approximate the solution to the respective MHE optimization problem at each time-step. Convergence is illustrated by a challenging small scale example, the Lorenz attractor with an unknown parameter. It is shown that the performance of the proposed real-time MHE algorithm is nearly identical to a fully converged MHE solution, while its fixed execution time per sample would allow one to solve 30 000 MHE problems per second on current hardware.

I. INTRODUCTION

Moving horizon estimation (MHE) is an optimization-based method for state estimation in which, at each time-step, only a finite window of measurement data is used to calculate the state estimate. MHE was developed [18], [22], [27], [28] to avoid the computational burden encountered by full-information estimators. This so-called “curse of dimensionality” arises from the need to solve optimization problems of ever increasing dimension as more measurements become available. Although restricted by this computational difficulty, full-information estimators have well developed convergence properties [22], [26], i.e., under certain conditions the estimated state is guaranteed to converge asymptotically to the true state of the system. A fundamental question, therefore, is to ask what convergence properties still hold for a given MHE scheme. Reviews of moving horizon estimation and its relation to other nonlinear filtering techniques can be found in [9], [25], [26].

It is shown in [22] that a MHE scheme for an unconstrained linear system can be defined which is equivalent to the classical Kalman Filter, and therefore inherits its well known stability properties. Convergence analysis of a simpler MHE implementation for linear unconstrained systems is given in [2], while constrained linear systems are discussed in [22], [23]. For nonlinear systems, various analyses of MHE convergence exist in the literature [3], [4], [16], [24], [26], [31]. These are discussed below in the context of real-time implementable moving horizon estimation.

For linear systems, the optimization required at each step of the MHE process is simply a quadratic program, and hence the associated computational cost is well understood. However, for nonlinear systems it is more difficult to bound the computational time needed to perform each optimization. Despite this, a common assumption in MHE convergence analysis [3], [23], [24] is that each optimization problem can be solved (either optimally or suboptimally) in the interval between sampling times. Unless this computational speed can be guaranteed for a particular problem, convergence of the resulting MHE scheme cannot be assured.

To avoid this issue, we consider a moving horizon scheme in which the optimization performed at each time-step is explicitly defined, meaning that the associated computational time has a calculable upper bound. The proposed scheme is defined formally in Section II, while an outline is given here. We consider a discrete time nonlinear system

\[
x_{k+1} = f_k(x_k) + v_k \\
y_k = h(x_k) + \eta_k
\]  

with additive measurement and state disturbances. The proposed MHE scheme is based on solving the optimization problem

\[
\min_{\xi_{(k-N:k)}} \frac{1}{2} \sum_{i=k-N}^k \|y_i - h(\xi_i)\|^2 \\
\text{s.t.} \quad \xi_{i+1} - f_i(\xi_i) = 0, \quad i = k - N, \ldots, k - 1
\]

at each time-step \(k\) using an iterative procedure. Given an initial guess \(\xi = \xi_{0k}\) for an optimal solution to (2) at time-step \(k\), an iterative method \(I_k\) is applied to approximate \(I_k(\xi_{0k}) := \xi_{k}^{(1)}, I_k(\xi_{k}^{(1)}) := \xi_{k}^{(2)}, \ldots\) to provide a sequence of vectors satisfying \(\xi_{k}^{(n)} \to \xi_{k}^{*}\), where \(\xi_{k}^{*}\) is an optimal solution to (2). However, instead of calculating the entire sequence, only the first iterate \(\xi_{k}^{(1)} = I_k(\xi_{0k}) := \{\xi_{1k}, \ldots, \xi_{Nk}\}\) is calculated and the state estimate \(\hat{x}_k\) at time \(t = k\) is given by \(\hat{x}_k := \xi_{k}^{(1)}\). Subsequently, \(\xi_{k}^{(1)}\) is shifted by a map \(S_k\) and used as the initialization of the iterative scheme at the next time-step \(t = k + 1\): \(\xi_{k+1}^{(1)} := S_k(\xi_{k}^{(1)}))\). It should be noted that in contrast to most MHE implementations, there is no prior weighting term in (2). Instead, information is passed between successive steps via the shifting initialization.

Several examples of ‘fast MHE’ schemes have recently been proposed. The most similar approach to the one analyzed in this paper is given in [16], but there are two important differences. First, the optimization problem (2) has a larger state space than the associated problem in [16], depending upon all elements \(\{\xi_{-N}, \ldots, \xi_{0}\}\) of the horizon window rather than just the first state \(\xi_{-N}\). This so-called simultaneous formulation complicates the convergence analysis presented in this paper, but is crucial for numerical stability and applicability to strongly nonlinear systems, as highlighted by the numerical example in Section III. Second, no generic convergence assumptions of the form \(\|I_k(x) - x_{k+1}\| \leq \alpha_k\|x - x_{k}\|\) will be made upon the iterative scheme in this paper. Instead, crucial analysis of a specific Gauss-Newton iterative scheme will be presented.

In [3], stability is considered for an ‘\(\epsilon\)-suboptimal’ form of MHE. In this approach, it is assumed that at each time-step decision variables are found for which the MHE objective function has a value within \(\epsilon\) of its optimal value. However, no specific numerical scheme is proposed to achieve this and, for example, if an iterative method is used, it may not be possible to guarantee \(\epsilon\)-suboptimality within a fixed number of iterations. Hence, this scheme cannot be guaranteed to work in real-time.

In [4], [31], the MHE scheme is accelerated by solving a background MHE problem with time intervals. Specifically, at time \(t = k - 1\) the current state estimate \(\hat{x}_{k-1}\) is used to provide a guess \(\hat{y}_k := f_k(\hat{x}_{k-1})\) for the measurement data at time \(k\). It is assumed that MHE optimization with data \(\{y_{k-N}, \ldots, y_{k-1}, \hat{y}_k\}\) can then be solved in the time interval \([k-1, k]\) to provide an initial state estimate \(\hat{x}_{k}^{(0)}\). When the true data \(y_k\) becomes available, the final state estimate is given by an update formula \(\hat{x}_{k} := \hat{x}_{k}^{(0)} + K_k(y_k - \hat{y}_k)\), where \(K_k\) is a matrix defined in terms of the background MHE problem. The point is that the calculation of the update can be performed extremely quickly. However, this technique still requires a background non-linear optimization to be solved within a fixed time interval. Again, even if it is proposed that this optimization is to be solved suboptimally, no guarantee can be given for computational speed unless a specific numerical scheme is considered.

For these reasons, we study a moving horizon estimation scheme based upon a one-step generalized Gauss-Newton procedure with shifting. In this way the computational burden at each time-step is fixed, meaning that the scheme presented can potentially be implemented in real-time. Observer design using numerical methods was originally proposed in [19], [21] for noise-free discrete time nonlinear systems, and for continuous time systems in [32]. In the discrete-time

¹Department of Aeronautics, Imperial College London, UK. Email: a.wynn@imperial.ac.uk
²Electrical Engineering Department (ESAT), K.U. Leuven, Belgium. Email: {milan.vukov,moritz.diehl} @esat.kuleuven.be
setting of [19], [21], the state estimate at time \( t = k \) is obtained by searching for a solution to \( \mathcal{Y}_k - H_k(\xi) = 0 \) by applying a fixed number of steps of Newton’s Algorithm. Here, \( \mathcal{Y}_k := [y_k - y_k^0 \cdots y_k^N] \), \( H_k(\xi) := [h(f_k(\xi)) \cdots h((f_{k-N}(\xi)))]' \), and if \( \xi^* \) is the obtained root, then the state estimate is given by forward simulation \( \hat{x}_k = (f_{k-1} \cdots \cdot f_{k-N})(\xi^*) \). Note that this is a sequential or single shooting approach, in which only the first element of this horizon window is used as an optimization variable.

Importantly, we study a simultaneous or multiple shooting formulation of the estimation problem, i.e. all states in the horizon are kept as variables of the optimization problem and the resulting nonlinearly constrained least-squares problem (2) is solved by the generalized Gauss-Newton Method, as proposed in the context of parameter estimation by Bock and Schlöder [7], [8], [30]. The use of the generalized Gauss-Newton method in the simultaneous framework is in contrast to the sequential approach described above where Newton’s method [19], [21] and variants such as Broyden’s method [20], hybrid-Newton [5], [6] and pseudo-Newton methods [13] have been applied.

An advantage of the simultaneous approach is that it can deal reliably with unstable and nonlinear dynamical systems, and even chaotic differential equations [15]. Roughly speaking, the reason for keeping the states as optimization variables, rather than eliminating them via a forward simulation, is that it leads to a well conditioned estimation problem. This is in contrast to the possibility of an ill conditioned estimation problem arising from the use of forward simulation, as highlighted by the numerical example in Section III, where it is shown that the simultaneous approach is more robust to incorrect initialization than the sequential approach. More information concerning the theoretical advantage of a simultaneous approach to Newton-based optimization for scalar-valued problems can be found in [1]. Due to the good contraction properties of the simultaneous approach, we can afford to perform only one Generalized Gauss-Newton iteration of the MHE problem, as originally proposed and investigated numerically, but without theoretical underpinning in [17].

The main result of the paper, Theorem 3, provides conditions under which this MHE scheme is asymptotically stable. Here we follow the spirit of convergence guarantees for related real-time algorithms for nonlinear model predictive control [11], [12]. In Section III a challenging numerical example is presented to illustrate the favourable convergence properties of the algorithm. The example concerns a chaotic system, the Lorenz attractor, subject to parameter changes.

A. Problem formulation and notation

A multiple horizon estimation scheme will be considered for (1), where the system state evolves \( (x_k)_{k \geq 0} \subset X \subseteq \mathbb{R}^n \). It is assumed that \( h : \mathbb{R}^n \rightarrow \mathbb{R}^p \) and each \( f_k : \mathbb{R}^n \rightarrow \mathbb{R}^n \) belong to \( C^2 \), and that \( \sup_{x \in X} \| \partial h_j f_k \|_{L^\infty(X)} \leq k_f \prec \infty \) and \( \| \partial h_j h \|_{L^\infty(X)} \leq k_h \prec \infty \), for each \( i \in \{1,2\}, j \in \{1, \ldots, n\} \). In particular, each \( f_k \) is Lipschitz on \( X \) with constant \( k_f \) and \( h \) is Lipschitz with constant \( k_h \). Throughout this note, the relation \( x \preceq y \) implies existence of a constant \( c \geq 0 \), depending only upon \( k_f, k_h \) and the horizon length \( N \), for which \( x \preceq cy \). To avoid confusion with this notation, matrix definiteness is denoted by the standard inequality symbol: \( A \succeq 0 \) if and only if \( x'Ax \geq 0 \), all \( x \in \mathbb{R}^n \). Measurement and state disturbances are assumed to be bounded such that \( \max \{\| \eta_k \|, \| \nu_k \| \} \leq \nu \) for each \( k \in \mathbb{N} \), for some \( \nu > 0 \).

At each time \( k \in \mathbb{N} \) only a finite horizon of measurement data of length \( N \in \mathbb{N} \) is available. The aim is to recreate the state of the system, at each time \( k \geq N + 1 \), by solving the minimization problem (2). Throughout this paper, the characters \( x_i \) are reserved to denote the true state of the system (1). Typically, variables representing the estimated state, for example, the decision variables of (2), are denoted \( \xi \). To perform the minimization (2), consider the Lagrangian

\[
L_k(w) := \frac{1}{2} \sum_{i=1}^{k+1} \| y_i - h(\xi_j) \|^2 + \sum_{i=k+1}^{k+1} \lambda_{i+1}(\xi_{i+1} - f_i(\xi_i)).
\]

with constraint multipliers \( \lambda_i \in \mathbb{R}^n \) and data vector \( w := (\xi_{k-N} \lambda_{k-N+1} \cdots \lambda_{k} \xi_{k})' \). The gradient \( F_k \) of the \( L_k \) is

\[
F_k(w) := \partial L_k(w)/\partial w,
\]

and the Gauss-Newton KKT matrix, \( H_k(w) \), is formed by replacing the second order derivative terms \( \partial^2 L_k(w) / \partial^2 w \) by zero in the full KKT matrix.

We consider a version of the MHE scheme in [17] where only one Gauss-Newton iteration is performed at each time-step, then the result shifted to form the initialization at the next time-step. At time \( k \geq N \), suppose that an initial value for the data vector, \( w_k^{(0)} \), is given. A single Gauss-Newton step is performed:

\[
\Delta w_k^{(n)} := -M_k(w_k^{(n-1)})^{-1}F_k(w_k^{(n-1)}), \quad w_k^{(1)} := w_k^{(0)} + \Delta w_k^{(0)}
\]

where \( \Delta w_k^{(n)} := (\xi_k^{(n)})' (\xi_{k-N}^{(n)})' \cdots (\lambda_{k-N+1}^{(n)})' (\lambda_{k}^{(n)})' \). The state estimate at time \( t = k \) is given by \( \hat{x}_k := \xi_k^{(1)} \). Before time \( t = k + 1 \), the initial data vector \( w_k^{(0)} \) is prepared by shifting \( w_k^{(1)} \). In particular, \( w_{k+1}^{(0)} := S_k(w_k^{(1)}) \), where

\[
S_k(\xi_{k-N} \lambda_{k-N+1} \cdots \lambda_{k} \xi_{k}) := (\xi_{k-N+1} \lambda_{k-N+2} \cdots \lambda_{k} \xi_{k} | 0 \cdots f_k(\xi_k))'.
\]

The two components of the algorithm, the Gauss-Newton step and the shifting initialization, are represented schematically in Figure 1.
II. CONVERGENCE OF THE REAL-TIME ESTIMATION ALGORITHM

It is assumed that the system is \((N + 1)\)-step observable.

**Assumption 1.** There exist constants \(r_\delta, \delta > 0\) such that
\[
\sum_{i=k-N}^{k} \|h(\xi_i^{(\nu)}) - h(\xi_i^{(\nu)})\|_2^2 \geq \delta^2 \|\xi_{k-N}^{(\nu)} - \xi_{k-N}^{(\nu)}\|_2^2,
\]
for each \(k \geq N\) and any trajectories \(\xi_i^{(\nu)} = f_i(\xi_i^{(\nu)})\) satisfying \(\|\xi_{k-N}^{(\nu)} - x_{k-N}\|_2 < r_\delta, j = 1, 2\).

The importance of observability is that it implies regularity of the Gaussian-Newton Hessian \(M_k(\cdot)\) on a tube of the form
\[
D_k(\nu) := \{ w : \|w - w_k(\nu)\| \leq \nu \} \subset \mathbb{R}^{(2N+1)n}
\]
centred upon the zero-multiplier data vector \(w_k(\nu) := (x_{k-N}^1 0 x_{k-N+1}^1 \cdots 0 x_k^1)^\top\) which contains the true solution at times \(k - N \leq t \leq k\). Proofs of the results in the section are given in Appendix A.

**Theorem 1.** Suppose Assumption 1 holds. Then there exist \(\tilde{r}, \tilde{\nu} > 0\) such that whenever \(\nu < \tilde{r}\) and \(\nu < \tilde{\nu}\):
(i) \(\sup_{k \geq N} \|M_k(w)\| \leq 1 + \delta^{-2}\);
(ii) There exist constants \(\kappa, \omega > 0\) such that
\[
\|M_k(w)^{-1}(M_k(w + \Delta w) - H_k(w + \Delta w)) \Delta w\| \leq \kappa \|\Delta w\|, \tag{5}
\]
\[
\|M_k(w)^{-1}(M_k(w + \Delta w) - M_k(w)) \Delta w\| \leq \omega \|\Delta w\|^2, \tag{6}
\]
for any \(k \geq N\), \(w, w' \in D_k(2\nu), t \in [0, 1] \) and \(\Delta w := w' - w\);
(iii) The constants satisfy \(\kappa \leq (r + \nu)(1 + \delta^{-2})\) and \(\omega \leq 1 + \delta^{-2}\);
(iv) \(F_k(w) = 0\) has at most one solution in \(D_k(2\nu)\).

Conditions (5) and (6) are the standard affine invariant assumptions for convergence of Newton-type methods [10], with (5) quantifying the quality of approximation that \(M_k\) provides to the true KKT matrix and (6) concerning Lipschitz continuity of \(M_k\). It will be assumed in the remainder of this paper that \(r < \tilde{r}, \nu < \tilde{\nu}\) are such that (5) holds for \(\kappa < 1\). Note that by Theorem 1(iii) this is always possible by the fact that \(f_k, h\) are Lipschitz implies that \(\mu, \sigma\) exist and are finite. We now state the main result, that the real-time estimation algorithm is locally convergent.

**Theorem 2.** Suppose Assumption 1 holds and that
\[
\epsilon := \left(\frac{\kappa + \frac{\omega\nu}{2}}{2}\right) \left(1 + \frac{\mu}{\nu} + \frac{\sigma}{1 - \kappa}\right) < 1, \quad \varphi := \min \left\{ \frac{1 - \epsilon}{1 - \epsilon}, \frac{1}{(1 + (1 + \nu)^\kappa/2)} \right\} > 0.
\]
Then if \(\|\Delta u_k\| \leq \varphi (1 - \epsilon)\), \(\lim_{k \rightarrow \infty} \|\hat{e}_k - x_k\| \leq \nu(\alpha + \epsilon)/(1 - \epsilon^{-1})\). \tag{11}

The condition \(\epsilon < 1\) comes from each of the constants related to the Gauss-Newton convergence (i.e. \(\kappa, \nu, \sigma\)) and the shift procedure (i.e. \(\mu, \sigma\)) influence stability of the overall MHE algorithm. Furthermore, since Theorem 1(iii) implies \(\kappa \leq (r + \nu)(1 + \delta^{-2})\), it follows that \(\epsilon < 1\) holds if \(r + \nu\) is sufficiently small. In turn, such a choice of \(r\) determines the maximum acceptable initial step-size \(\Delta_{\text{max}}\). Note that, by continuity of \(w \rightarrow \Delta w\) and since \(\|\Delta u_k\| = 0\), this step-size condition is satisfied if the distance of the initial guess from the fully converged solution \(\|w_k^{(0)} - w_k^{(\nu)}\|\) is sufficiently small, i.e., a ‘good’ initial guess is required. Robustness of the MHE scheme to initialization is explored numerically in Section III. Finally, the constant \(\varphi\) influences the maximum acceptable noise magnitude via \(\nu < \varphi\) and, similarly, \(\varphi > 0\) holds if \(r + \nu\) is sufficiently small.

III. NUMERICAL CASE STUDY

In this section the one-step Gauss-Newton MHE algorithm is applied to a challenging nonlinear test problem, the augmented Lorenz attractor
\[
x^{(1)} = 10(x^{(2)} - x^{(1)}), \quad x^{(1)} = x^{(4)} x^{(1)} - x^{(2)} x^{(3)}
\]
\[
x^{(2)} = -28x^{(1)} x^{(2)} + x^{(3)}, \quad x^{(3)} = 0
\]
with initial condition \(x_0 := (-1 3 4 9)^\top\). The states \(x^{(1)}, x^{(2)}, x^{(3)}\) define standard Lorenz attractor equations, while the parameter \(p\) is assumed to be unknown and is estimated by introducing an auxiliary state \(x^{(0)}\). The system is discretized using a fourth-order Runge-Kutta scheme (RK4) with an integrator step of \(t_i = 0.01\) and five integrator steps per sampling interval, resulting in a discretization time-step of \(t = 5t_i = 0.05\). The time-independent function \(f\) in the system dynamics is defined as the result of those five steps of RK4 applied to the augmented Lorenz attractor. The measured output is the first state \(y_k := x_k^{(1)} \in \mathbb{R}^4\), i.e., \(h(x) = x^{(1)}\) and measurements are corrupted with disturbances drawn from a normal distribution with covariance \(\nu\). No state noise is added. However, the parameter \(p\) is assumed to make a large unphysical change in value, from 25 to 30, at time \(t = 1.5\). The MHE algorithm is initialized at time step \(k = N\) by setting the shooting node \(\xi_{k-N}\) to be an arbitrary initial guess \(\xi_{k-N}^{(0)} = f(\xi_{k-N-1}^{(0)})\), then initializing the remaining nodes \(\xi_{k-N}^{(0)} = f(\xi_{k-N-1}^{(0)})\) via forward simulation.

Performance is evaluated for each MHE simulation by calculating the average sum of squares error (SSE) statistic
\[
\text{SSE}(N, \nu) := \frac{1}{N_{\text{sim}}} \sum_{k=1}^{N_{\text{sim}}} \sum_{j=1}^{3} \left| x_k^{(j)} - x_k^{(0)} \right|^2.
\]
over \( N_{\text{sim}} = 200 \) timesteps, i.e. an interval of length \( 200t_d \). Each entry in Table I represents the mean SSE statistic for a given \((N, \nu)\) pair taken over 1000 instances of the noise at covariance level \( \nu \).

In each simulation, the MHE algorithm is initialized via forward simulation from \( \xi_0^{(0)} = (-2 \ 4.5 \ 20)^T \), while the true system starts at \( x_0 = (-1.3 \ 4.25)^T \). It can be observed from Table I that MHE performance improves as \( N \) increases and worsens as \( \nu \) increases. Furthermore, MHE performance is observed to be poor for the small horizon length \( N = 5 \), suggesting that observability is close to being lost at this horizon length. Interestingly, for \( \nu = 0, 10^{-3} \), SSE is higher for \( N = 15 \) than \( N = 10 \). This is explained by the fact that MHE with a longer horizon recovers more slowly from the jump in \( \rho \), as can be seen in Figure 2 (see \( 1.5 \leq t \leq 2.25 \)).

Fig. 2. MHE estimates for \( y_k := x_k^{(1)} , \nu = 0 \) and \( N = 10, N = 15 \).

Table II contains the averages of three data vector errors for the case \( N = 15 \). Here, \( d_1 := \frac{1}{N_{\text{sim}}} \sum_{k=1}^{N_{\text{sim}}} \| w_k^{(1)} - w_k^{(2)} \| \) is the average distance between \( w_k^{(1)} \) and the true data vector \( w_k^{(2)} \), while \( d_* := \frac{1}{N_{\text{sim}}} \sum_{k=1}^{N_{\text{sim}}} \| w_k^{(0)} - w_k^{(2)} \| \) is the average fully iterated data vector error. For comparison, \( d_1^w := \frac{1}{N_{\text{sim}}} \sum_{k=1}^{N_{\text{sim}}} \| w_k^{(1)} - w_k^{(0)} \| \) is the average data vector error associated with a one-step single-shooting MHE algorithm of the type analyzed in [21]. The statistics \( d_1, d_*, d_1^w \) stated in Table II are the averages of \( d_1, d_*, d_1^w \) taken over 1000 simulations at each noise covariance level \( \nu \). Associated mean SSE statistics are in brackets. In all simulations, each MHE algorithm is initialized via forward simulation from \( \xi_0^{(0)} = (-2 \ 4.5 \ 20)^T \).

The general trend \( d_* \leq d_1 \leq d_1^w \) is clear from Table II. This is exemplified by the particular case \( N = 15, \nu = 0.01 \) shown in Figure 3: the fully converged data vector \( d_* \) recovers faster than both \( d_1 \) and \( d_1^w \) from the incorrect initialization of the MHE scheme (see \( 0.75 \leq t \leq 1.5 \)); and both \( d_* \) and \( d_1 \) recover faster than \( d_1^w \) from the step change in \( \rho \) (see \( 2.25 \leq t \leq 2.5 \)). As expected, each of the mean data vector errors increase as \( \nu \) increases.

Table I

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>( N = 5 )</th>
<th>( N = 10 )</th>
<th>( N = 15 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu = 0 )</td>
<td>1.15 0.51 0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu = 10^{-3} )</td>
<td>14.5 0.54 0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu = 10^{-2} )</td>
<td>141 0.83 0.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu = 10^{-1} )</td>
<td>2230 3.90 1.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table II

<table>
<thead>
<tr>
<th>( \nu )</th>
<th>( d_1 )</th>
<th>( d_* )</th>
<th>( d_1^w )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \nu = 0 )</td>
<td>1.55 (0.54)</td>
<td>1.34 (0.34)</td>
<td>2.54 (1.56)</td>
</tr>
<tr>
<td>( \nu = 10^{-3} )</td>
<td>2.02 (0.55)</td>
<td>1.80 (0.34)</td>
<td>2.89 (1.56)</td>
</tr>
<tr>
<td>( \nu = 10^{-2} )</td>
<td>3.03 (0.61)</td>
<td>2.83 (0.40)</td>
<td>3.86 (1.70)</td>
</tr>
<tr>
<td>( \nu = 10^{-1} )</td>
<td>6.31 (1.19)</td>
<td>6.15 (0.98)</td>
<td>7.28 (3.40)</td>
</tr>
</tbody>
</table>

Fig. 3. MHE performance with \( y_k := x_k^{(1)}, N = 15 \) and \( \nu = 0.01 \). Data vector errors for multiple shooting (dashed line), single shooting (dotted) and fully converged (solid grey) MHE implementations.

Fig. 4. Percentage of successful MHE runs for one-step multiple shooting (MS) and single shooting (SS) schemes when initialized at a distance \( R \) from the initial condition \( x_0 = (-1.3 \ 4.25)^T \).

To assess MHE stability with respect to incorrect initialization, both the one-step single shooting (SS) and multiple shooting (MS) MHE schemes are simulated from perturbed initializations

\[ \xi_0^{(0)} = (-1 \ 3 \ 4.25)^T + (\xi_0 \ 0 \ 0)^T, \quad \xi_0 \in \mathbb{R}^3, \quad \| \xi_0 \|_2 = R. \]

The parameter \( \rho \) is assumed to jump from 25 to 30 at \( t = 1.5 \), only the first state is observed, and, for all simulations, zero measurement noise is assumed. At each perturbation radius \( R \), 100 randomly generated perturbations \( \xi_0 \) with \( \| \xi_0 \|_2 = R = 25 \) were simulated and both SS and MS MHE applied from the respective perturbed initializations. The percentages of successful MHE runs are displayed in Figure 4, where a simulation is deemed unsuccessful if the condition number of the Gauss-Newton Hessian \( M_k(w) \) is bigger than \( 10^{10} \) at any given time-step.

1Here, single shooting refers to the method described in Section I in which data vectors are defined by forward simulating a single state optimization variable over the given horizon window. Specifically, in the notation described in Section I, at time-step \( k \) one Gauss-Newton step is applied to search for a solution to \( y_k - H_k(\xi) = 0 \) from the initial guess \( \xi = \xi_0^{(0)} \).

2Average statistics for SS MHE in the case \( \nu = 10^{-1} \), marked by 1, are only taken over successful MHE runs (condition number of KKT matrix less than \( 10^{10} \) for all timesteps). In this case only 64% of runs were successful.
IV. SUMMARY

Conditions are proven for local convergence of a real-time MHE scheme based upon performing one Gauss-Newton iteration followed by a shifting procedure. The algorithm considered is simultaneous (all states in the measurement window are optimizable variables), which is shown by means of a numerical study involving the Lorenz attractor to outperform a single shooting methodology both in terms of convergence rate and robustness to initialization. Investigation of the theoretical reasons behind this improvement will form the basis of future research.

APPENDIX A

Proof of Theorem 1: First define the observability matrix
\[ O(w) := \left\{ \frac{\partial h}{\partial \xi}(\xi_{k-N})^\top, \frac{\partial h}{\partial \xi}(\xi_{k-N+1})A_{k-N}(\xi_{k-N})^\top, \ldots, \frac{\partial h}{\partial \xi}(\xi_{k-1})A_{k-1}(\xi_{k-1}) \cdots A_{k-N}(\xi_{k-N})^\top \right\}^\top \]
where \( A_i := \frac{\partial h}{\partial w} \) and \( w := (\xi_{k-N}, \lambda_{N-k-N} \cdots \lambda_k \xi_k)^\top \). Then, up to second order terms, the function
\[ \xi_{k-N} \mapsto \frac{1}{2} \sum_{i=k-N}^k \| h(\xi_i) - h(z_i) \|^2, \]
where \( \xi_{i+1} := f_i(\xi_i), z_{k-N} := x_{k-N} \) and \( z_{i+1} := f_i(z_i) \) has Taylor expansion about the true state \( x_{k-N} \) equal to
\[ (\xi_{k-N} - x_{k-N})^\top O(\tilde{w}_{k-N}^\top)O(\tilde{w}_{k-N}^\top)(\xi_{k-N} - x_{k-N}). \]
Here, \( \tilde{w}_{k-N}^\top := (z_{k-N}^\top z_{k-N+1}^\top \cdots 0^\top 0^\top)^\top \). Assumption 1 implies that
\[ O(\tilde{w}_{k-N}^\top)O(\tilde{w}_{k-N}^\top) \geq \delta^2 I. \] (12)
Now, since each \( f_k, h \) are Lipschitz, \( \| u_{k-N}^\top - \tilde{w}_{k-N}^\top \| \leq \nu \) and \( w \mapsto O(w)O(w) \) is Lipschitz. It follows that (12) exists at \( t > \bar{t} \) such that \( O(\tilde{w}_{k-N}^\top)O(\tilde{w}_{k-N}^\top) \geq \delta^2 I/2 \) whenever \( 0 \leq \nu < \nu_3 \). Again, using Lipschitz continuity of \( w \mapsto O(w)O(w) \), there exists a constant \( \bar{r} > 0 \) such that \( O(w)O(w) \geq \delta^2 I/4 \) whenever \( w \in D_k(2r) \) and \( r < \bar{r} \).

We now show that \( M_k := \sup\{ \| M_k(w)^{-1} \| : w \in D_k(2r), k \geq N \} \leq 1 + \delta^{-2} \). First, let \( y \in \mathbb{R}^{(2N+1)n} \) and \( w \in D_k(2r) \). We aim to find a vector \( x := (x_1, \ldots, x_{2N+1})^\top \), \( x_i \in \mathbb{R}^n \) such that \( M_k(w)x = y \). Using the definition of \( M_k(w) \), it can be shown that such an \( x \) exists if and only if \( x_1 \) is a solution to the equation \( F(\tilde{w}_{k-N}) \tilde{w}_{k-N} x_1 = Fy \) where \( F : \mathbb{R}^{(2N+1)n} \rightarrow \mathbb{R}^n \) is a linear map satisfying \( \| F \| \leq 1 \). Moreover, if \( x_1 \) is unique then \( x := (x_1, \ldots, x_{2N+1})^\top \) is a unique solution of \( M_k(w)x = y \) and each \( x_i \) can be expressed as a linear transformation of \( x_1 \) and \( y \). Since \( O(w)O(w) \) is invertible on \( D_k(2r) \), there does exist such a unique \( x_1 \in \mathbb{R}^n \) and, furthermore, \( \| x_1 \| \leq \| F \| \| y \| \delta^{-2} \), from which \( M_k \leq 1 + \delta^{-2} \) follows. That \( M_k \) is independent of \( k \geq N \) follows since \( k, l, k_h \) are independent of \( k \).

To show (5), let \( w, w' \in D_k(2r), \Delta w = w' - w \) and \( t \in [0, 1] \).

Proof of Theorem 2: That the Newton iterates converge to a point \( w_{k-N}^\ast \in D_k(2r) \) is a standard result (see, e.g., [12, Thm. 2]). The following inequalities which are derived in such a proof will be required subsequently: for each \( i \geq 0 \)
\[ \| \Delta w_{k-N}^{i+1} \| \leq \left( \kappa + \frac{\nu}{2} \right) \| \Delta w_{k-N}^i \|, \]
and
\[ \| \Delta w_{k-N}^{i} - \Delta w_{k-N}^0 \| \leq \left( 1 - \kappa - \frac{\nu^2}{4} \right) \| \Delta w_{k-N}^0 \|^{-1} \| \Delta w_{k-N}^i \| \leq \left( \frac{\kappa + \nu^2}{1 - \kappa} \right) \| \Delta w_{k-N}^0 \|. \] (13)

To prove the bound on the fully converged data vector, let \( w_{k-N}^\ast = (\xi_{k-N}^\ast \lambda_{k-N}^\ast \cdots \lambda^\ast \xi^\ast) \) and note that
\[ \| \Delta w_{k-N}^\ast - \Delta w_{k-N}^0 \|^2 = \sum_{i=k-N+1}^k \| \lambda_i^\ast \|^2 + \sum_{i=k-N}^k \| \xi_i^\ast - x_i \|^2. \] (15)

Now, since \( F_k(w_{k-N}^\ast) = 0 \), and the Lipschitz continuity of \( f, h \) implies
\[ \sum_{i=k-N+1}^k \| \lambda_i^\ast \|^2 \leq \nu^2 + \sum_{i=k-N}^k \| \xi_i^\ast - x_i \|^2 \leq \nu^2 + \| \xi_{k-N}^\ast - x_{k-N} \|^2. \]
It is therefore sufficient to suitably bound \( \| \xi_{k-N}^\ast - x_{k-N} \| \). Observability (using \( w_{k-N}^\ast \in D_k(2r) \) and \( 2r < 2\tilde{r} \leq r_3 \)) and optimality imply
\[ \delta^2 \| \xi_{k-N}^\ast - x_{k-N} \|^2 \leq \sum_{i=k-N}^k \| h(\xi_i^\ast) - h(z_i) \|^2 \leq 4 \sum_{i=k-N}^k \| h(\xi_i - y_i) \|^2, \]
where \( z_{k-N} := x_{k-N} \). The result follows since \( f_i \) and \( h \) Lipschitz implies that \( \| h(z_i - y_i) \| \leq \nu \), for each \( i \).

Proof of Theorem 3: Assume initially that \( w_{k-N}^0 \in D_N(r) \) and \( \| \Delta w_{k-N}^0 \| \leq \Delta_{\max} \). We first show that \( w_{k-N+1}^0 \in D_{k-N+1}(r) \). Theorem 2 implies that there exists \( w_{k-N+1}^0 \in D_{k-N+1}(r) \) for which \( w_{k-N}^0 \rightarrow w_{k-N}^\ast \).

3 All simulations performed on a standard PC with a quad-core Intel Q9650 processor, running 64-bit Ubuntu Linux 10.10.
Furthermore, (7), (13), (14) and \( \| \Delta w^{(0)}_N \| \leq \Delta_{\text{max}} \) imply

\[
\| w^{(1)}_N - w^{(r)}_N \| \leq \| w^{(1)}_N - w^{(r)}_N \| + \| w^{(r)}_N - w^{(0)}_N \| \\
\leq \tau \left( \frac{\tau + \sqrt{2}}{1 - \sqrt{2}} \right) + \alpha \nu.
\]

Now, \( \nu \leq \varphi \) gives \( w^{(1)}_N \in D_N(r/2(1 + 2\kappa^2_1)^2) \subset D_N(r^2 - 2\nu^2)/2 \) where the final inclusion follows since \( \nu < r/2 \). The definition of \( S_N \) and the fact that \( f_i \) are Lipschitz imply that \( w^{(0)}_{N+1} = S_N(w^{(1)}_N) \in S_N(D_N(r^2 - 2\nu^2)/2(1 + 2\kappa^2_1)^2) \subset D_{N+1}(r) \).

Since \( \| \Delta w^{(0)}_N \| \leq \Delta_{\text{max}} \). Since \( w^{(1)}_N \in S_N^{-1}(D_{N+1}(r)) \cap D_N(2\tau) \), (8), (9) and (10) give

\[
\| \Delta w^{(0)}_{k+N} \| \leq \mu \| M_N(w^{(1)}_N) \|^{-1} F_N(w^{(1)}_N) \| + \sigma \| w^{(1)}_N - w^{(r)}_N \| + \nu
\]

(by (4) and (7)) \( \leq \mu \| w^{(1)}_N - w^{(r)}_N \| + \sigma \| w^{(1)}_N - w^{(0)}_N \| + \sigma(1 + \alpha) \nu \)

(by (13) and (14)) \( \leq \epsilon \| \Delta w^{(0)}_N \| + \sigma(1 + \alpha) \nu \).

Since \( \nu \leq \varphi \leq (1 - \epsilon) \Delta_{\text{max}}/(1 + \alpha) \sigma \) it follows that \( \| \Delta w^{(0)}_{k+N} \| \leq \Delta_{\text{max}} \).

We may now apply (16) recursively to obtain

\[
\limsup_{k \to \infty} \| \Delta w^{(0)}_{k+N} \| \leq \nu(1 + \alpha) \sigma (1 - \epsilon)^{-1}. \tag{17}
\]

Inequalities (7) and (13), (14) imply

\[
\| w^{(1)}_{k+N} - w^{(r)}_{k+N} \| \leq \| w^{(1)}_{k+N} - w^{(r)}_{k+N} \| + \alpha \nu
\]

\[
\leq \left( \frac{\tau + \sqrt{2}}{1 - \sqrt{2}} \right) \| \Delta w^{(0)}_{k+N} \| + \alpha \nu.
\]

The bound (11) now follows from (17), the definition of \( \epsilon \) and the fact that \( \| x_{k+N} - x_{k+N} \| \leq \| x_{k+N} - x_{k+N} \| \). Finally, (13), (14) and the actual assumption \( | \Delta w^{(0)}_N | \leq \| (1 - \epsilon) \Delta_{\text{max}} \| \leq \| w^{(1)}_N - w^{(r)}_N \| \leq \alpha \nu + r(1 - 1/\sqrt{2}) \leq \alpha \nu + r(1 - 1/\sqrt{2}) \leq r \), where the final inequality follows since, trivially, \( \varphi \leq r/\sqrt{2}\alpha \). Hence, \( w^{(0)}_N \in D_N(r) \), which completes the proof.

\[\square\]

ACKNOWLEDGEMENTS

The research was supported by the Research Council KUL, via the Optimization in Engineering Center OPTEC (CoE EF05/006 and PFV/10/002), GOA/11/05 Ambiorics, GOA/10/09 MaNet, IOF-SCORES4CHEM and PhD/postdoc/fellow grants, the Flemish Government via FWO (PhD/postdoc grants, projects GO226.06, GO321.06, GO302.07, GO320.08, GO558.08, GO557.08, GO588.09, GO377.09, research communities ICCoS, ANMM, MLD) and via IWT (PhD Grants, Eureka-Flite+, SBO LeCoPro, SBO Cimaqs, SBO POM, O&O-Dsquare), the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011), the IBBT, Contract Research (AMINAL), viCERP, ACCM, and the EU via FP7-EMBOCON (ICT-248940), FP7-HD-MPC (INFSoS-ICT-223854), ERNSI, COST interlICIS, FP7-SADC (MC ITN-264735), as well as the ERC project HIGHWIND (259 166).

REFERENCES