Zero-Order Moving Horizon Estimation

Katrin Baumgärtner, Andrea Zanelli, Moritz Diehl

***Abstract*— Moving Horizon Estimation (MHE) is an optimization-based approach to nonlinear state estimation. The computational burden associated with the online solution of the corresponding nonlinear optimization problems poses a major challenge when applying MHE in practice. Motivated by these considerations, we introduce zero-order MHE, an inexact, but computationally less expensive variant of exact MHE. Zero-order MHE is based on the Gauss-Newton algorithm and avoids online evaluation of derivatives and factorizations. As for exact MHE, the estimation error produced by zero-order MHE would become zero in the absence of noise and model-plant mismatch, and grows linearly with the noise level. In addition, we present a structure-exploiting approach for recursive factorization of the Gauss-Newton Hessian approximation which allows for efficient arrival cost updates. Zero-order MHE is compared to exact and linear MHE both theoretically, in terms of estimation error bounds, and numerically, by applying the methods to a state estimation example.**

I. Introduction

Moving Horizon Estimation (MHE) is an optimization-based approach to nonlinear state estimation where the state estimate is determined by solving a nonlinear optimization problem that takes into account a fixed number of measurements on a moving horizon in the past.

Although previous work, [1], [2], has shown that MHE might outperform the widely used Extended Kalman Filter (EKF), the application of MHE is in practice often prevented by its considerably higher computational complexity [1], [3].

Under the assumption of Gaussian noise and in the absence of inequality constraints, the MHE state estimate is obtained as the solution of an unconstrained nonlinear least-squares problem. This kind of optimization problem is usually stated as a nonlinear root-finding problem such that a local minimizer can be obtained via Newton-type iterations. The computational cost of an MHE iteration is thus mainly associated with sensitivity generation and factorization of the Hessian approximation used within the Newton-type iterations.

In this paper, we present and analyse zero-order MHE, an inexact, but computationally less expensive variant of MHE, which is based on the Gauss-Newton (GN) algorithm. The main idea is to use fixed derivatives to avoid the computational cost induced by online sensitivity generation and factorization. The method is a direct translation of the ideas proposed in [4], [5] from the context of Nonlinear Model Predictive Control (NMPC) to the case of MHE.

Furthermore, we present an efficient method for updating the arrival cost within zero-order MHE which is based on a recursive factorization of the Hessian approximation.

In practice, we usually have to satisfy strict upper bounds on the computation time that is available per iteration. In this case, the number of Newton-type iterations has to be limited which directly leads to the Real-Time Iteration (RTI) framework [6], [4], [7], [8]. In addition to exact and zero-order MHE, we therefore apply the corresponding RTI variants to an example state estimation problem and discuss the estimation results.

The paper is structured as follows. The exact MHE formulation as well as linear and zero-order MHE are introduced in Section II. In Section III, we present upper bounds on the estimation error produced by exact, linear and zero-order MHE. In Section IV, the recursive factorization approach for the Hessian approximation is described. In Section V, a multi-level algorithm that combines zero-order MHE and the real-time iteration framework is outlined. In Section VI, we apply exact, zero-order and linear MHE as well as the corresponding RTI variants to a textbook example and discuss the estimation results, which show competitive performance of zero-order MHE.

II. Moving Horizon Estimation Formulation And Algorithms

In the following, the Moving Horizon Estimation (MHE) formulation and zero-order MHE are introduced. Furthermore, we briefly describe linear MHE which will serve as a baseline in both Section III and Section VI.

For notational simplicity, we regard time-invariant systems of the following form:

$$x_{i+1} = f(x_i) + w_i, \tag{1a}$$

$$y_i = h(x_i) + v_i, \tag{1b}$$

with states $x_i \in \mathbb{R}^{n_x}$, measurements $y_i \in \mathbb{R}^{n_y}$, state and measurement disturbances $w_i \in \mathbb{R}^{n_x}$ and $v_i \in \mathbb{R}^{n_y}$. The disturbances w_i, v_i are assumed to follow a zero-mean Gaussian distribution. We make the assumption that the functions f and h are twice continuously differentiable.

We consider the following MHE formulation:

$$\begin{aligned} \min_{x} \quad & \frac{1}{2}\|x_0 - \hat{x}_0\|^2_{Q_0} \; + \; \frac{1}{2}\sum_{i=0}^{N-1} \|x_{i+1} - f(x_i)\|^2_Q \\ & + \; \frac{1}{2}\sum_{i=0}^{N} \|h(x_i) - y_i\|^2_R, \end{aligned} \tag{2}$$

This research was supported by DFG via Research Unit FOR 2401, and by the German Federal Ministry for Economic Affairs and Energy (BMWi) via eco4wind (0324125B) and DyConPV (0324166B).

Katrin Baumgärtner and Andrea Zanelli are with the Department of Microsystems Engineering (IMTEK) and Moritz Diehl is with the Department of Microsystems Engineering (IMTEK) and Department of Mathematics, University Freiburg, 79110 Freiburg, Germany.

katrin.baumgaertner@pluto.uni-freiburg.de

with optimization variables $x = (x_0, \ldots, x_N)$. The matrices $Q_0, Q \in \mathbb{R}^{n_x \times n_x}$ and $R \in \mathbb{R}^{n_y \times n_y}$ are assumed to be symmetric and positive definite.

Let $V(x)$ be the objective function of the above optimization problem. It can be rewritten in the following compact form:

$$V(x) = \frac{1}{2}\|M(x) - Y\|^2, \tag{3}$$

where

$$M(x) = W^{\frac{1}{2}} \begin{bmatrix} x_0 \\ x_1 - f(x_0) \\ \vdots \\ x_N - f(x_{N-1}) \\ h(x_0) \\ \vdots \\ h(x_N) \end{bmatrix}, \; Y = W^{\frac{1}{2}} \begin{bmatrix} \hat{x}_0 \\ 0 \\ \vdots \\ 0 \\ y_0 \\ \vdots \\ y_N \end{bmatrix}, \tag{4}$$

and $W = \text{diag}(Q_0, Q, \ldots, Q, R, \ldots, R)$. Thus, the MHE problem corresponds to solving the following nonlinear least-squares problem:

$$\min_x \; \frac{1}{2} \|r(x)\|^2 \tag{5}$$

with residual function $r(x) = M(x) - Y$.

A. Exact MHE and the Gauss-Newton Algorithm

If the MHE problem (5) is solved with the Gauss-Newton (GN) algorithm and $x^k = (x_0^k, \ldots, x_N^k)$ denotes the current iterate, the next GN iterate x^{k+1} is given by the equation

$$B(x^k)\left(x^{k+1} - x^k\right) = -J\left(x^k\right)^\top r\left(x^k\right), \tag{6}$$

where $J\left(x^k\right) = \frac{\partial r}{\partial x}\left(x^k\right) = \frac{\partial M}{\partial x}\left(x^k\right)$ is the Jacobian of the residual function and $B(x^k) = J\left(x^k\right)^\top J\left(x^k\right)$ is the GN Hessian approximation.

Within each GN iteration, a substantial part of the overall computation time is required for evaluation of $J\left(x^k\right)$ and factorization of the GN Hessian approximation $B\left(x^k\right)$.

B. Zero-order MHE

Zero-order MHE does not evaluate derivatives online, but instead uses a fixed Jacobian approximation $\bar{J}$ and, in particular, a fixed Hessian approximation $\bar{B} = \bar{J}^\top \bar{J}$ which can be evaluated and factorized offline. The method is called zero-order, as only function evaluations and no higher order derivatives are required online.

The zero-order GN iteration is defined by the following equation:

$$\bar{B}\left(x^{k+1} - x^k\right) = -\bar{J}^\top r\left(x^k\right), \tag{7}$$

where $\bar{B} = \bar{J}^\top \bar{J}$ is the fixed Hessian approximation, x^k is the current iterate, x^{k+1} is the subsequent iterate.

The computational cost of a single zero-order GN step corresponds to the evaluation of the residual function $r\left(x^k\right)$ and a linear system solve with prefactorized Hessian approximation $\bar{B}$.

Comparing (7) with the original GN iteration (6), where the computation of a single step requires evaluation of $r\left(x^k\right)$ and $J\left(x^k\right)$ as well as factorization of $B\left(x^k\right)$, the computational cost is significantly reduced.

Assuming that the zero-order GN iteration converges, the zero-order MHE estimate is defined as the limit of the sequence of zero-order GN iterates.

C. Linear MHE

With linear MHE, the nonlinear model is approximated by a fixed linearization, which is used across all MHE iterations. In this case, the optimization problem that has to be solved in each iteration reduces to a convex, quadratic program (QP):

$$x^{\text{lin}} \; \in \; \arg\min_x \; \frac{1}{2} \left\|M_{\text{lin}}(x, \bar{x}) - Y\right\|^2, \tag{8}$$

where $\bar{x}$ denotes the linearization point. The linearized model M_{lin} is given by

$$M_{\text{lin}}(x, \bar{x}) = M(\bar{x}) + \bar{J}(x - \bar{x}) \tag{9}$$

with $\bar{J} = J(\bar{x})$. Assuming that $\bar{B} = \bar{J}^\top \bar{J}$ is invertible, the solution x^{lin} to (8) is unique and given by

$$x^{\text{lin}} = \bar{x} - \bar{B}^{-1}\bar{J}^\top \left(M(\bar{x}) - Y\right). \tag{10}$$

D. Asymptotic Contraction Rate

Local convergence of the zero-order MHE scheme follows directly from the convergence analysis in [4].

If the exact and zero-order GN iterates converge to x^{exact} and x^{zo} respectively, the asymptotic contraction rate [9] is given by

$$\kappa^{\text{exact}} = \rho\left(\mathbb{I} - B\left(x^{\text{exact}}\right)^{-1} \frac{\partial^2 V}{\partial x^2}\left(x^{\text{exact}}\right)\right), \tag{11}$$

$$\kappa^{\text{zo}} = \rho\left(\mathbb{I} - \bar{B}^{-1}\left(\bar{J}^\top J(x^{\text{zo}})\right)\right), \tag{12}$$

where $\rho(\cdot)$ denotes the spectral radius. If $\bar{J}$ is similar to $J(x^{\text{zo}})$, we would thus expect fast (local) convergence of zero-order MHE.

III. Estimation Error Bounds

In the following, we assess the estimation error, which is the deviation of the estimated state trajectory from the true state trajectory, as a function of the measurement noise. We make the simplifying assumption that the true state trajectory is not disturbed by noise.

Let $x^{\text{true}} = (x_0^{\text{true}}, \ldots, x_N^{\text{true}})$ denote the true state trajectory where, by assumption,

$$x_{k+1}^{\text{true}} = f(x_k^{\text{true}}), \quad k = 0, \ldots, N-1.$$

We define the disturbance vector $\epsilon \in \mathbb{R}^{n_\epsilon}$, $n_\epsilon = (N+1)(n_x + n_y)$, as

$$\epsilon = W^{\frac{1}{2}}\left(\hat{x}_0 - x_0^{\text{true}}, w_0, \ldots, w_{N-1}, v_0, \ldots, v_N\right),$$

where $w_k = 0$, $k = 0, \ldots, N$ - 1, as we assumed that there is no state noise. We can then define Y as a function of ϵ:

$$Y(\epsilon) = M(x^{\text{true}}) + \epsilon. \tag{13}$$

More explicitly, equation (13) is given as

$$W^{\frac{1}{2}} \begin{bmatrix} \hat{x}_0 \\ 0 \\ \vdots \\ 0 \\ y_0 \\ \vdots \\ y_N \end{bmatrix} = W^{\frac{1}{2}} \begin{bmatrix} x_0^{\text{true}} + (\hat{x}_0 - x_0^{\text{true}}) \\ x_1^{\text{true}} - f(x_0^{\text{true}}) \\ \vdots \\ x_N^{\text{true}} - f(x_{N-1}^{\text{true}}) \\ h(x_0^{\text{true}}) + v_0 \\ \vdots \\ h(x_N^{\text{true}}) + v_N \end{bmatrix}.$$

Consequently, the residual function r is parametric in ϵ,

$$r(x, \epsilon) = M(x) - Y(\epsilon), \tag{14}$$

with $M(x)$ as defined in (4) and $Y(\epsilon)$ as defined in (13).
In order to derive bounds on the estimation error, we regard the disturbance vector ϵ as a parameter within the optimization problems that are solved by exact, zero-order and linear MHE.
If the exact GN iterates converge to $x^{\text{exact}}(\epsilon)$ for given $Y(\epsilon)$, then $x^{\text{exact}}(\epsilon)$ is a solution to the parametric nonlinear root-finding problem defined by:

$$J(x)^\top r(x, \epsilon) = 0. \tag{15}$$

With $x^k = x^{k+1}$, this follows directly from (6). Equation (15) corresponds to the first-order necessary condition of optimality of the original MHE formulation (2) given as:

$$P_{\text{exact}}(\epsilon): \quad x^{\text{exact}}(\epsilon) \in \arg\min_x \; V(x, \epsilon), \tag{16}$$

where $V(x, \epsilon) = \frac{1}{2}\|r(x, \epsilon)\|^2$, and $r(x, \epsilon)$ as defined in (14).
If the zero-order GN iterates converge to $x^{\text{zo}}(\epsilon)$, then $x^{\text{zo}}(\epsilon)$ is a solution to the parametric nonlinear root-finding problem defined by:

$$\bar{J}^\top r(x, \epsilon) = 0, \tag{17}$$

which is directly obtained from equation (7) by setting $x^{k+1} = x^k$. As shown in [4], this equation can be interpreted as the first-order condition of optimality of a perturbed variant of the original problem $P_{\text{exact}}(\epsilon)$, which is given by:

$$P_{\text{zo}}(\epsilon, \bar{x}): \quad x^{\text{zo}}(\epsilon, \bar{x}) \in \arg\min_x \; V_{\text{zo}}(x, \epsilon, \bar{x}), \tag{18}$$

where

$$V_{\text{zo}}(x, \epsilon, \bar{x}) = \frac{1}{2}\|r(x, \epsilon)\|^2 + g(\epsilon, \bar{x})^\top x,$$
$$g(\epsilon, \bar{x}) = \left(\bar{J} - J(x^{\text{zo}}(\epsilon, \bar{x}))\right)^\top r(x^{\text{zo}}(\epsilon, \bar{x}), \epsilon),$$

with $\bar{J} = J(\bar{x})$. For linear MHE, the corresponding parametric QP is given as

$$P_{\text{lin}}(\epsilon, \bar{x}): \quad x^{\text{lin}}(\epsilon, \bar{x}) \in \arg\min_x \; V_{\text{lin}}(x, \epsilon, \bar{x}), \tag{19}$$

where

$$V_{\text{lin}}(x, \epsilon, \bar{x}) = \frac{1}{2}\|r_{\text{lin}}(x, \epsilon, \bar{x})\|^2,$$
$$r_{\text{lin}}(x, \epsilon, \bar{x}) = M_{\text{lin}}(x, \bar{x}) - Y(\epsilon),$$

with $M_{\text{lin}}(x, \bar{x})$ as in (9) and $Y(\epsilon)$ as in (13).

Lemma 1. *Suppose $J(x^{\text{true}})^\top J(x^{\text{true}})$ is invertible. Under this assumption, there is a nonempty neighbourhood $\mathcal{N}(x^{\text{true}})$ such that both $J(x)^\top J(x^{\text{true}})$ and $J(x)^\top J(x)$ are invertible for all $x \in \mathcal{N}(x^{\text{true}})$.*

Proof. The eigenvalues of a matrix are given as the roots of its characteristic polynomial. The coefficients of the characteristic polynomial are continuous functions of the entries of the matrix. As the roots of a polynomial function depend continuously on its coefficients [10], we can conclude that the eigenvalues of a matrix are a continuous function of its entries.
Continuity of $J(x)$ thus implies continuity of the eigenvalues of $J(x)^\top J(x^{\text{true}})$ and $J(x)^\top J(x)$ with respect to x. As $J(x^{\text{true}})^\top J(x^{\text{true}})$ has only nonzero eigenvalues, there must be a nonempty neighbourhood $\mathcal{N}(x^{\text{true}})$ of x^{true} such that the eigenvalues of $J(x)^\top J(x^{\text{true}})$ and $J(x)^\top J(x)$ are nonzero for all $x \in \mathcal{N}(x^{\text{true}})$. □

Theorem 1. *Suppose $J(x^{\text{true}})^\top J(x^{\text{true}})$ is invertible. Then there are nonempty neighbourhoods $\mathcal{N}(0) \subseteq \mathbb{R}^{n_\epsilon}$ and $\mathcal{N}(x^{\text{true}}) \subseteq \mathbb{R}^{(N+1)n_x}$ of zero and the true state trajectory such that the solution maps*

$$x^{\text{exact}} : \mathcal{N}(0) \to \mathbb{R}^{n_x},$$
$$x^{\text{zo}} : \mathcal{N}(0) \times \mathcal{N}(x^{\text{true}}) \to \mathbb{R}^{n_x},$$
$$x^{\text{lin}} : \mathcal{N}(0) \times \mathcal{N}(x^{\text{true}}) \to \mathbb{R}^{n_x},$$

with $x^{\text{exact}}(0) = x^{\text{zo}}(0, \bar{x}) = x^{\text{lin}}(0, x^{\text{true}}) = x^{\text{true}}$, are well-defined and the following holds:

(a) $\left\|x^{\text{true}} - x^{\text{exact}}(\epsilon)\right\| = O(\|\epsilon\|)$,
(b) $\left\|x^{\text{true}} - x^{\text{zo}}(\epsilon, \bar{x})\right\| = O(\|\epsilon\|)$,
(c) $\left\|x^{\text{true}} - x^{\text{lin}}(\epsilon, \bar{x})\right\| = O(\|\epsilon\|) + O\left(\|x^{\text{true}} - \bar{x}\|^2\right)$.

Proof. For part (a), consider again the nonlinear root-finding problem defined by

$$J(x)^\top r(x, \epsilon) = 0. \tag{20}$$

If there is no noise, the residuals at x^{true} are zero, i.e. $r(x^{\text{true}}, 0) = 0$. Thus, x^{true} satisfies (20) for $\epsilon = 0$. With $r(x^{\text{true}}, 0) = 0$, we additionally have

$$\frac{\partial}{\partial x}\left(J(x)^\top r(x, \epsilon)\right)\Big|_{(x,\epsilon)=(x^{\text{true}},0)} = J(x^{\text{true}})^\top J(x^{\text{true}}).$$

By assumption $J(x^{\text{true}})^\top J(x^{\text{true}})$ is invertible. We can therefore apply the Implicit Function Theorem at $(x, \epsilon) = (0, x^{\text{true}})$, which implies that the solution to (20) is locally unique, i.e. the solution map $x^{\text{exact}}(\cdot)$ is well-defined. Besides, the solution map satisfies $x^{\text{exact}}(0) = x^{\text{true}}$ and is Lipschitz continuous in a neighbourhood of zero.

For part (b) and (c), we first note that Lemma 1 implies that there is a nonempty neighbourhood $\mathcal{N}(x^{\text{true}})$ such that $J(\bar{x})^\top J(x^{\text{true}})$ and $J(\bar{x})^\top J(\bar{x})$ are invertible for all $\bar{x} \in \mathcal{N}(x^{\text{true}})$. For both linear and zero-order MHE, we will consider only linearization points within this neighbourhood.

For part (b), suppose $\bar{x} \in \mathcal{N}(x^{\text{true}})$. We consider the nonlinear root-finding problem given by

$$\bar{J}^\top r(x, \epsilon) = 0, \tag{21}$$

with $\bar{J} = J(\bar{x})$. For $\epsilon = 0$, the true state trajectory is a solution to (21), since $r(x^{\text{true}}, 0) = 0$. Besides, the Jacobian

$$\frac{\partial}{\partial x}\left(\bar{J}^\top r(x,\epsilon)\right)\Big|_{(x,\epsilon)=(x^{\text{true}},0)} = \bar{J}^\top J(x^{\text{true}})$$

is invertible by assumption. Therefore, we can apply the Implicit Function Theorem at $(x, \epsilon) = (x^{\text{true}}, 0)$. It implies that the solution to (21) is locally unique and can thus be described by $x^{\text{zo}}(\,\cdot\,, \bar{x})$. The solution maps satisifies $x^{\text{zo}}(0, \bar{x}) = x^{\text{true}}$ and is Lipschitz continuous w.r.t. ϵ in a neighbourhood of zero.

For part (c), we again consider $\bar{x}$ in $\mathcal{N}(x^{\text{true}})$ such that $J(\bar{x})^\top J(\bar{x})$ is invertible. In this case, $x^{\text{lin}}(\epsilon, \bar{x})$ is explicitly given as:

$$x^{\text{lin}}(\epsilon, \bar{x}) = \bar{x} - \bar{B}^{-1}\bar{J}^\top \left(M(\bar{x}) - Y(\epsilon)\right). \tag{22}$$

The solution map $x^{\text{lin}}(\,\cdot\,, \bar{x})$ is continuously differentiable in ϵ which implies that

$$x^{\text{lin}}(\epsilon, \bar{x}) = x^{\text{lin}}(0, \bar{x}) + O\left(\|\epsilon\|\right). \tag{23}$$

We now consider $x^{\text{lin}}(0, \bar{x},)$, which is continuously differentiable in $\bar{x}$. At $\bar{x} = x^{\text{true}}$, we have $M(\bar{x}) - Y(0) = 0$. Therefore, the derivative at $\bar{x} = x^{\text{true}}$ is given as

$$\frac{\partial x^{\text{lin}}}{\partial \bar{x}}(0, \bar{x})\Big|_{\bar{x}=x^{\text{true}}} = \mathbb{I} - \bar{B}^{-1}\bar{J}^\top\bar{J}\Big|_{\bar{x}=x^{\text{true}}} = 0.$$

As the first-order derivative is zero at $\bar{x} = x^{\text{true}}$, the Taylor expansion of $x^{\text{lin}}(0, \bar{x})$ at $\bar{x} = x^{\text{true}}$ is given by

$$x^{\text{lin}}(0, \bar{x}) = x^{\text{lin}}(0, x^{\text{true}}) + O\left(\|\bar{x} - x^{\text{true}}\|^2\right).$$

Plugging this into (23), we obtain

$$x^{\text{lin}}\left(\epsilon, \bar{x}\right) = x^{\text{lin}}\left(0, x^{\text{true}}\right) + O\left(\|\bar{x} \text{ - } x^{\text{true}}\|^2\right) + O\left(\|\epsilon\|\right).$$

As $x^{\text{lin}}(0, x^{\text{true}}) = x^{\text{true}}$, we can conclude that

$$\|x^{\text{lin}}(\epsilon, \bar{x}) - x^{\text{true}}\| = O\left(\|\bar{x} - x^{\text{true}}\|^2\right) + O\left(\|\epsilon\|\right).$$

□

IV. Recursive Factorization and Efficient Arrival Cost Updates

Up to now, we assumed that the only quantity that changes from one MHE problem to the next is the vector Y which includes the measurements $y_0, \ldots, y_N$. Within the MHE framework, we would, however, update the arrival cost as well. In the following, we present a recursive approach for the factorization of the Hessian approximation $\bar{B}$ that allows for efficient updates of the arrival cost.

Let $\bar{J} = J\left(\bar{x}\right)$ where $\bar{x} = (\bar{x}_0, \ldots, \bar{x}_0)$. We assume that $\bar{J}$ has full rank, which implies that $\bar{B} = \bar{J}^\top\bar{J}$ is positive definite.

The linear system, which has to be solved in each zero-order GN iteration, is given as:

$$\bar{J}^\top\bar{J}\; p = c, \tag{24}$$

where $p = x^{k+1} - x^k$ denotes the GN step. The right-hand side c is given as $c = -\bar{J}^\top r\left(x^k\right)$.

To simplify notation, we consider the case $N = 3$. The linear system (24) is then explicitly given as

$$\begin{bmatrix} P_0 & (QA)^\top & & \\ QA & P_1 & (QA)^\top & \\ & QA & P_2 & (QA)^\top \\ & & QA & P_3 \end{bmatrix}\begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}, \tag{25}$$

where $A = -\frac{\partial f}{\partial x}(\bar{x}_0)$, $C = \frac{\partial h}{\partial x}(\bar{x}_0)$ and

$$\begin{aligned} P_0 &= Q_0 + A^\top QA + C^\top RC, \\ P_k &= Q + A^\top QA + C^\top RC, \quad k = 1, 2, \\ P_3 &= Q + C^\top RC. \end{aligned}$$

We can exploit the block-tridiagonal structure of the matrix $\bar{B} = \bar{J}^\top\bar{J}$ to recursively eliminate variables from (25) using Schur complements. Let $\bar{P}_3 = P_3$ and $\bar{c}_3 = c_3$. The linear system that is defined by the last two equations of (25) then reads:

$$\begin{bmatrix} QA & P_2 & (QA)^\top \\ & QA & \bar{P}_3 \end{bmatrix}\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} c_2 \\ \bar{c}_3 \end{bmatrix}. \tag{26}$$

As $\bar{B}$ is positive definite, we have $\bar{P}_3 \succ 0$. The last equation of (26) thus implies

$$p_3 = \bar{P}_3^{-1}\left(\bar{c}_3 - QAp_2\right).$$

Together with the first equation of (26), we obtain

$$QAp_1 + \underbrace{\left(P_2 - A^\top Q^\top\bar{P}_3^{-1}QA\right)}_{\bar{P}_2} p_2 = \underbrace{c_2 - A^\top Q^\top\bar{P}_3^{-1}\bar{c}_3}_{\bar{c}_2}.$$

The system of linear equations (25) that we originally started with has been reduced to

$$\begin{bmatrix} P_0 & (QA)^\top & \\ QA & P_1 & (QA)^\top \\ & QA & \bar{P}_2 \end{bmatrix}\begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ \bar{c}_2 \end{bmatrix}. \tag{27}$$

This reduced matrix is again positive definite, as it is the Schur complement $\bar{B}/\bar{P}_3$ of $\bar{P}_3$ in $\bar{B}$. Thus, the reduced linear system (27) has exactly the same structure as the original linear system (24) such that we can proceed eliminating variables by applying the procedure we just described.

In summary, the recursion is given by the following formulas.

Backward sweep:

$$\begin{aligned} \bar{P}_N &= P_N, \\ \bar{P}_k &= P_k - A^\top Q^\top\bar{P}_{k+1}^{-1}QA, \quad k = N-1, \ldots, 0, \\ \bar{c}_N &= c_N, \\ \bar{c}_k &= c_k - A^\top Q^\top\bar{P}_{k+1}^{-1}\bar{c}_{k+1}, \quad k = N-1, \ldots, 0. \end{aligned}$$

Forward sweep:

$$p_0 = \bar{P}_0^{-1}\bar{c}_0,$$
$$p_{k+1} = \bar{P}_{k+1}^{-1}\left(\bar{c}_{k+1} - QAp_k\right), \qquad k = 0, \ldots, N-1.$$

With this method, the computational cost associated with the factorization of the Hessian approximation is $O\left(Nn_x^3\right)$. For an overview of alternative block-structured factorization approaches for MHE, we refer to [11].

If we want to update P_0, which is the only quantity depending on the arrival cost matrix Q_0, the last step of the recursion has to be recomputed. This is associated with a computational cost of $O\left(n_x^3\right)$. Within zero-order MHE, it is thus possible to efficiently update the arrival cost online, while avoiding a full refactorization of the Hessian approximation.

V. Multi-Level Real-Time Iteration Algorithm

In the following, we outline a multi-level real-time iteration algorithm similar to the method proposed in [4]. The main idea of the Real-Time Iteration (RTI) framework, as introduced in [6], [4], is to perform only a single Newton-type step per iteration, thereby significantly reducing the computational cost. Previous work, both in the context of NMPC [6], [4] and MHE [7], has shown that RTI methods achieve competitive performance.

Zero-order MHE, as presented in Section II, does not require any evaluation of derivatives. It might be nevertheless beneficial to update the Hessian approximation from time to time, especially, if the considered system is operated at a wide range of different operating points. In this case, we propose to combine a real-time iteration variant of zero-order MHE with a higher level process that updates the Hessian approximation and its factorization.

At the low level, we compute, at every timestep k, an estimate of the current state of the system by performing a single zero-order GN step. The higher level process, that works in parallel, takes the current estimate as input and uses it as linearization point for updating the Hessian approximation. The new Hessian is factorized, which might require multiple timesteps, and the factorization is provided to the lower level process as soon as it is available. A schematic presentation of this algorithm is given in Fig. 1.

An update of the Hessian approximation might be initiated after a fixed number of timesteps, i.e. every N_u-th timestep, or even asynchronously, e.g. when a change of the steady state is detected.

VI. Illustrative Example

In this section, we apply exact, linear, and zero-order MHE as well as the corresponding real-time iteration variants to an example problem in order to illustrate the algorithms and results presented in the previous sections.

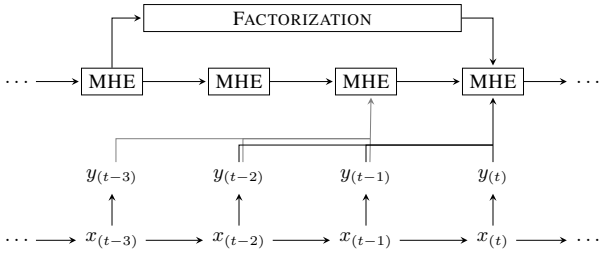

Fig. 1. Schematic representation of the multi-level algorithm with horizon $N = 2$. We use indices in parentheses to distinguish them from the indices used within the MHE formulation.

The following example has been adopted from [12]. We consider a continuous stirred-tank reactor where an irreversible, first-order reaction, $A \rightarrow B$, occurs. The continuous time system is given as

$$\dot{T} = \frac{F_0(T_0 - T)}{\pi r^2 h} - \frac{\Delta H k_0 c}{\rho C_p}\exp\left(-\frac{E}{RT}\right) + \frac{2Uh\left(T_\mathrm{c} - T\right)}{r\rho C_p},$$
$$\dot{c} = \frac{F_0(c_0 - c)}{\pi r^2 h} - k_0 c \exp\left(-\frac{E}{RT}\right),$$
$$\dot{T}_\mathrm{c} = 0,$$

where c is the molar concentration of the substance A, T is the reactor temperature, and T_c is the coolant liquid temperature. The values of the model parameters can be found in [12]. Only the reactor temperature T can be measured. We discretize the continuous time system with $\Delta t = 0.25$ min using a single RK4 step.

The system has a stable steady state at

$$x_{\mathrm{s}_1} = (324.497,\ 877.825,\ 300),$$

which we use as linearization point for linear MHE. In addition, the derivative at x_{s_1} is used as Jacobian approximation within zero-order MHE, i.e. we set $\bar{J} = J(\bar{x})$ with $\bar{x} = (x_{\mathrm{s}_1}, \ldots, x_{\mathrm{s}_1})$. For all estimation methods, the horizon is set to $N = 10$. As weighting matrices, we use $Q = \mathrm{diag}\left(10, 10, 10^6\right)$ and $R = 0.1$. The initial arrival cost is set to $Q_0 = \mathrm{diag}\left(100, 10, 1\right)$ and $\hat{x}_0 = x_{\mathrm{s}_1}$.

For updating the arrival cost, we apply an Extended Kalman Filter (EKF) update and prediction step. More precisely, if $\hat{x}_0$ and Q_0 define the current arrival cost, we obtain the updated arrival cost quantities, $\hat{x}_0^+$ and Q_0^+, from the following formulas:

Update step:

$$\hat{x}_0^- = \hat{x}_0 + K\left(y_0 - h(\hat{x}_0)\right),$$
$$Q_0^- = Q_0(\mathbb{I} - KC)^{-1},$$
$$K = Q_0^{-1}C^\top\left(CQ_0^{-1}C^\top + R^{-1}\right)^{-1}.$$

Predict step:

$$Q_0^+ = \left(A(Q_0^-)^{-1}A^\top + \tilde{Q}^{-1}\right)^{-1},$$
$$\hat{x}_0^+ = f\left(\hat{x}_0^-\right),$$

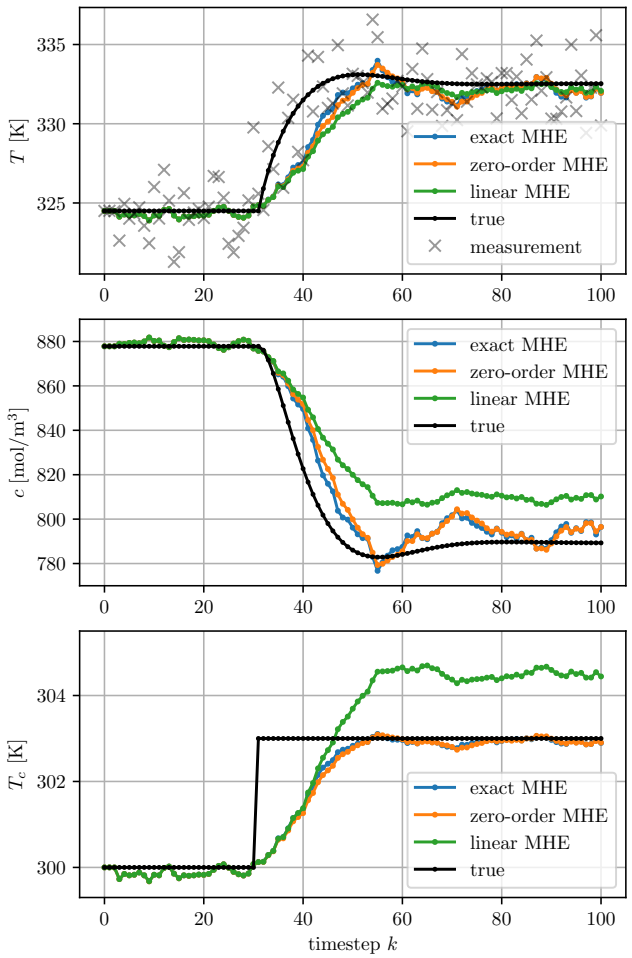

Fig. 2. True state trajectory, measurements and MHE estimates obtained from exact, zero-order and linear MHE. The horizon is $N = 10$.

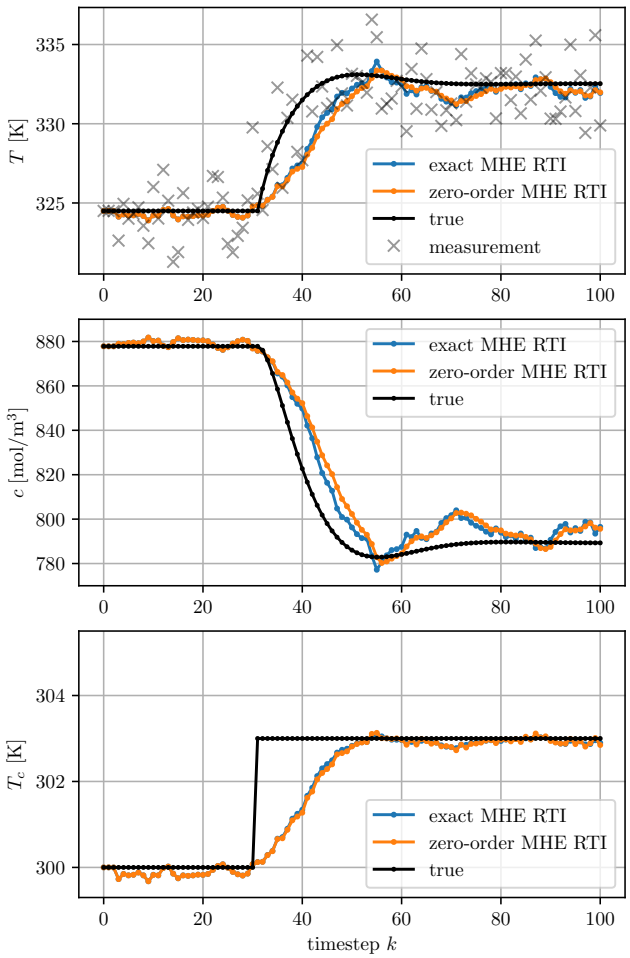

Fig. 3. True state trajectory, measurements and MHE estimates obtained from the RTI variants of exact and zero-order MHE. The horizon is $N = 10$.

where we do not use the original weighting matrix Q, which is used within the MHE formulation, but the matrix $\tilde{Q}$. By choosing $\tilde{Q}$ very small, we can decrease the influence of the arrival cost, and therefore the influence of measurements outside of the horizon, on the overall cost function. Here, we choose $\tilde{Q} = \text{diag}(10, 10, 10)$ in order to allow the coolant liquid temperature T_c, which is assumed to be constant on the horizon, to change its value from one MHE iteration to the next.

For exact MHE, the derivatives $A = \frac{\partial f}{\partial x}(\hat{x}_0^-)$ and $C = \frac{\partial h}{\partial x}(\hat{x}_0)$ are used for updating the arrival cost. For zero-order MHE, we use the fixed derivatives that are used within the zero-order GN iterations, i.e. $A = \frac{\partial f}{\partial x}(x_{\text{s}_1})$, $C = \frac{\partial h}{\partial x}(x_{\text{s}_1})$, such that the arrival cost update does not require sensitivity generation.

As intialization, we use a shifted version of the solution that we obtained in the previous iteration plus a one-step ahead prediction. More precisely, if $(x_0^*, \dots, x_N^*)$ is the solution of the previous MHE problem, we initialize the subsequent problem with $(x_1^*, \dots, x_N^*, f(x_N^*))$.

Fig. 2 shows the true state trajectory and the observed data, as well as the state estimates obtained from the exact, linear and zero-order MHE. As illustrated by Fig. 2, the system is initially in the steady state x_{s_1}. There is no state noise except for a step change of the coolant liquid temperature T_c occuring at timestep $k = 30$ which serves as an unmodeled disturbance and changes the steady state of the system to $x_{\text{s}_2} = (332.5,\ 782.2,\ 303)$.

All three estimation approaches lead to almost identical estimates around the steady state x_{s_1}, which was used as linearization point. After the step change, only exact MHE and zero-order MHE recover the new steady state x_{s_2}. The state estimates obtained from linear MHE are reasonably accurate only for the reactor temperature T. For the states that are not measured, i.e. concentration c and coolant liquid temperature T_c, the linear MHE estimates deviate significantly from the true state trajectory.

In Fig. 3, the estimates obtained from the RTI variant of exact and zero-order MHE are plotted. For both methods, a single GN step is sufficient to achieve results comparable to the fully converged solutions of exact MHE.

Fig. 4 shows the asymptotic contraction rates κ^{exact} and κ^{zo} as defined in (11) and (12) computed at the corresponding solutions shown in Fig. 2. The contraction rate for zero-order

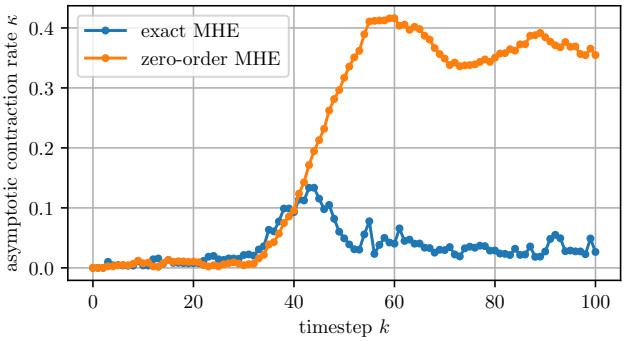

Fig. 4. Asymptotic contraction rate κ for zero-order and exact MHE.

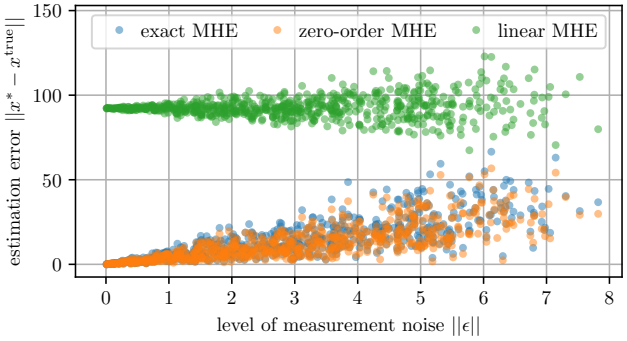

Fig. 5. Estimation error as a function of $\|\epsilon\|$.

MHE increases after the step change of the coolant liquid temperature T_c, as we move away from the linearization point. The contraction rate for exact MHE increases after the step change, when the residuals are high, and then decreases again, as the new steady state of the system is estimated correctly again.

Fig. 5 serves as an illustration of Theorem 1. To obtain the depicted data, we generate random measurement noise trajectories and solve the corresponding MHE problems using exact, linear and zero-order MHE. We use a horizon of $N = 10$ and the same weighting matrices as before. As linearization point, the steady state x_{s_1} is used, the true state of the system is however the steady state x_{s_2}.

VII. Conclusions And Outlook

For exact MHE, a considerable computational burden is associated with sensitivity generation and factorization of the Hessian approximation which has to be performed online. We propose an inexact zero-order method that avoids these costs by using fixed derivatives and thus a fixed Hessian approximation throughout the estimation process. Compared to the standard GN algorithm, the zero-order variant reduces the computational complexity of a single GN iteration significantly while producing an estimation error which is of the same order as the estimation error of exact MHE. The even simpler approach of using a fixed linearization, i.e. linear MHE, yields the same order of estimation error only if we linearize at the true state of the system.

Due to the special structure of the Hessian approximation, we can obtain its factorization using a backward recursion procedure which allows for an efficient online update of the arrival cost within zero-order MHE.

Additionally, we outline a multi-level algorithm that combines a real-time iteration variant of zero-order MHE with asynchronous updates of the Hessian approximation.

The effectiveness of zero-order MHE in comparison to exact and linear MHE is shown numerically using a state estimation example. For the presented example, the corresponding real-time iteration variant of zero-order MHE achieves an estimation performance similar to the performance of exact MHE.

In this paper, we focus on the unconstrained case only. However, previous work has shown that the algorithm can easily be extended to the constrained case and will then always yield a feasible solution [13]. Therefore, future work will extend the presented results to constrained MHE and, additionally, include an efficient implementation of the presented algorithm as well as extensive benchmarking.

References

[1] E. Haseltine and J. Rawlings, "Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation," *Industrial and Engineering Chemistry Research*, vol. 44, pp. 2451–2460, 2005.

[2] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246–258, 2003.

[3] J. B. Rawlings and B. R. Bakshi, "Particle filtering and moving horizon estimation," *Computers and Chemical Engineering*, vol. 30, pp. 1529–1541, 2006.

[4] H. G. Bock, M. Diehl, E. A. Kostina, and J. P. Schlöder, "Constrained optimal feedback control of systems governed by large differential algebraic equations," in *Real-Time and Online PDE-Constrained Optimization*. SIAM, 2007, pp. 3–22.

[5] A. Zanelli, R. Quirynen, and M. Diehl, "An efficient inexact NMPC scheme with stability and feasibility guarantees," in *Proceedings of 10th IFAC Symposium on Nonlinear Control Systems*, Monterey, CA, USA, August 2016.

[6] M. Diehl, *Real-Time Optimization for Large Scale Nonlinear Processes*, ser. Fortschritt-Berichte VDI Reihe 8, Meß-, Steuerungs- und Regelungstechnik. Düsseldorf: VDI Verlag, 2002, vol. 920, phD Thesis.

[7] P. Kühl, M. Diehl, T. Kraus, J. P. Schlöder, and H. G. Bock, "A real-time algorithm for moving horizon state and parameter estimation," *Computers and Chemical Engineering*, vol. 35, no. 1, pp. 71–83, 2011.

[8] A. Alessandri and M. Gaggero, "Fast moving horizon state estimation for discrete-time systems using single and multi iteration descent methods," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4499–451, 2017.

[9] M. Diehl, *Lecture Notes on Numerical Optimization*, 2016, (Available online: http://cdn.syscop.de/publications/Diehl2016.pdf).

[10] R. Naulin and C. Pabst, "The roots of a polynomial depend continuously on its coefficients," *Rev. Colombiana Mat*, vol. 28, pp. 35–37, 1994.

[11] G. Frison, "Algorithms and methods for high-performance model predictive control," Ph.D. dissertation, Technical University of Denmark (DTU), 2015.

[12] G. Pannocchia and J. Rawlings, "Disturbance Models for Offset-Free Model-Predictive Control," *AIChE Journal*, vol. 49, pp. 426–437, 2003.

[13] A. Zanelli, R. Quirynen, and M. Diehl, "Efficient zero-order NMPC with feasibility and stability guarantees," in *Proceedings of the European Control Conference (ECC)*, Naples, Italy, June 2019.