2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Moving Horizon Estimation with Adaptive Regularization for Ill-Posed State and Parameter Estimation Problems

Katrin Baumgärtner, Rudolf Reiter, Moritz Diehl

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9993416

***Abstract*— We investigate the usage of Moving Horizon Estimation (MHE) for state and parameter estimation for partially non-detectable systems with measurements corrupted by outliers. We propose an arrival cost update formula based on the Generalized Gauss-Newton method and illustrate how it can be generalized to nonconvex loss functions that can be effectively used for outlier rejection. Moreover, we propose an adaptive regularization scheme for the arrival cost which introduces *forgetting* as well as additional pseudo-measurements to the arrival cost update. We illustrate the performance of the proposed algorithms on a longitudinal vehicle state and parameter estimation problem.**

I. Introduction

Moving Horizon Estimation (MHE) is an optimization-based method for nonlinear state and parameter estimation. The estimates are obtained as the solution of a nonlinear optimization problem that takes into account a fixed number of previous observations [1]–[5]. The contributions of observations outside of the *horizon* are summarized by an – often quadratic – arrival cost term. As the optimization window is shifted from one iteration to the next, the arrival cost term is typically updated to account for the measurement dropping out of the optimization window.

The formulation of the estimation problem in terms of a nonlinear program allows us to explicitly account for non-Gaussian noise distributions by choosing appropriate loss functions. For outlier rejection, nonconvex loss function are of particular interest [6]. Moreover, additional system parameters can be easily estimated alongside the state by considering an augmented system model [3], [7]. If the states and parameters are estimated online as part of a closed-loop control system, the augmented model might however no longer be detectable as soon as a steady-state is reached. As a consequence, the covariance associated with the state and parameter estimates grows unboundedly leading to high sensitivity to outlier measurements.

The contribution of this paper is twofold: On the one hand, we propose a general update formula for the arrival cost for general convex and – in some special cases – even nonconvex loss functions. The update approach generalizes the method proposed in [8] which considered only the L_2-loss. On the other hand, we illustrate how the arrival cost can be regularized when additional parameters are estimated in order to bound the associated covariance, thus reducing the sensitivity to outliers. The regularization introduces *forgetting*, i.e. a downweighting of past measurements, as well as additional pseudo-measurements of the estimated parameters within the arrival cost update. We illustrate the accuracy of the proposed MHE formulation on a longitudinal vehicle motion estimation problem. In contrast to [9], where MHE is applied to a similar problem, we refrain from using a Pacejka-type *Magic Formula* [10] and consider a simple model for the tire transmission force parameterized by only two parameters, which has been used in [11].

II. Model Assumptions and MHE Formulation

In the following, we introduce the augmented system model including additional system parameters that are to be estimated. Besides, the Moving Horizon Estimation (MHE) problem is defined.

We regard discrete-time systems subject to additive state and measurement disturbances,

$$z_{i+1} = \tilde{f}(z_i, u_i, \theta) + \tilde{w}_i, \tag{1a}$$
$$y_i = \tilde{h}(z_i) + v_i, \tag{1b}$$

with states $z_i \in \mathbb{R}^{n_z}$, control inputs $u_i \in \mathbb{R}^{n_u}$, measurements $y_i \in \mathbb{R}^{n_y}$, as well as state and measurement disturbances $w_i \in \mathbb{R}^{n_x}$ and $v_i \in \mathbb{R}^{n_y}$. The model equations depend on some unknown parameters $\theta \in \mathbb{R}^{n_\theta}$. The functions $\tilde{f}$ and $\tilde{h}$ are assumed to be twice continuously differentiable. In order to estimate the parameters θ alongside the system state z_i, we will use the following augmented system formulation

$$x_{i+1} = f(x_i, u_i) + w_i, \tag{2a}$$
$$y_i = h(x_i) + v_i, \tag{2b}$$

with

$$x_i := \begin{bmatrix} z_i \\ \theta_i \end{bmatrix}, \quad f(x_i, u_i) := \begin{bmatrix} \tilde{f}(z_i, u_i, \theta_i) \\ \theta_i \end{bmatrix}, \quad h(x_i) = \tilde{h}(z_i).$$

With this formulation, the parameters θ_i might change from one time step to the next. The magnitude of this change in parameters can be chosen by selecting an appropriate loss function. The inputs u_i are assumed to be known exactly, we thus introduce the shorthand $f_i(x_i) := f(x_i, u_i)$.

At time step $m = n + N$, the MHE formulation with measurement window $y = (y_n, y_{n+1} \ldots, y_m)$ is given as:

$$\min_{x_n, \ldots, x_m} \alpha_n(x_n) + \sum_{i=n}^{m-1} \psi(x_{i+1} - f_i(x_i)) + \sum_{i=n}^{m} \phi(h(x_i) - y_i) \tag{3}$$

with optimization variables $(x_n, \ldots, x_m)$. We assume that $\psi : \mathbb{R}^{n_x} \to \mathbb{R}_+$ and $\phi : \mathbb{R}^{n_y} \to \mathbb{R}_+$ are continuously

This research was supported by DFG via Research Unit FOR 2401 and project 424107692 and by the EU via ELO-X 953348.

Katrin Baumgärtner and Rudolf Reiter are with the Department of Microsystems Engineering (IMTEK), University of Freiburg. Moritz Diehl is with the Department of Microsystems Engineering (IMTEK) and Department of Mathematics, University of Freiburg.
katrin.baumgaertner@imtek.uni-freiburg.de

differentiable. The arrival cost $\alpha_n : \mathbb{R}^{n_x} \to \mathbb{R}$, is assumed to be twice continuously differentiable. We denote the solution of problem (3) by $\left(\hat{x}_{n|m}, \hat{x}_{n+1|m}, \ldots, \hat{x}_{m|m}\right)$.

The first term within the cost function in (3) represents the arrival cost approximation $\alpha_n(x_n)$. This term should approximate the exact arrival cost [4], which we denote by $\alpha_n^*(x_n)$. The exact arrival cost is defined via the recursion

$$\alpha_{i+1}^*(x_{i+1}) = \min_{x_i} \alpha_i^*(x_i) + \psi(x_{i+1} - f_i(x_i)) + \phi(h(x_i) - y_i).$$

If the exact arrival cost $\alpha_n^*(x_n)$ is used instead of $\alpha_n(x_n)$ within the MHE problem (3), we would recover the same solution as the full information problem (FIE) [4], as can be easily shown by dynamic programming arguments. As the exact arrival cost $\alpha_n^*(x_n)$ is generally intractable to compute, it is often replaced by a quadratic approximation, here denoted by $\alpha_n(x_n)$,

$$\alpha_n(x_n) = \frac{1}{2}\|x_n - \breve{x}_n\|^2_{P_n^{-1}}. \tag{4}$$

The arrival cost mean $\breve{x}_n$ and covariance matrix P_n are typically updated from one MHE problem to the next in order to incorporate the contribution of the measurement dropping out of the MHE horizon.

III. Arrival Cost Update

In this section, we derive a general formula for updating a quadratic arrival cost term by approximating the exact update at the current state estimate. Note that the following approach is applicable for any system of the form given in (2) and might also be used if only states and no system parameters are estimated. A similar derivation for the special case of quadratic loss functions $\psi(w_n)$ and $\phi(v_n)$ is given in [8].

When updating the arrival cost from one iteration to the next, we would ideally like to approximate the exact arrival cost update, which is – for a given arrival cost $\alpha_n(x_n)$ – defined as

$$\tilde{\alpha}_{n+1}(x_{n+1}) = \min_{x_n} \underbrace{\alpha_n(x_n) + \psi(x_{n+1} - f_n(x_n)) + \phi(h(x_n) - y_n)}_{=:\tilde{V}_n(x_n, x_{n+1})} \tag{5}$$

We approximate $\tilde{\alpha}_{n+1}$ by solving a quadratic approximation of (5), given as:

$$\alpha_{n+1}(x_{n+1}) = \min_{x_n} \quad \tilde{V}_n^{\text{QUAD}}(x_n, x_{n+1}) \tag{6}$$

where $\tilde{V}_n^{\text{QUAD}}$ is a quadratic approximation of $\tilde{V}_n$ and is assumed to be of the following form:

$$\begin{aligned} \tilde{V}_n^{\text{QUAD}}(x_n, x_{n+1}) = & \\ & \frac{1}{2}\begin{bmatrix}\Delta x_n \\ \Delta x_{n+1}\end{bmatrix}^\top B_n(\hat{x}_{n|m}, \hat{x}_{n+1|m}) \begin{bmatrix}\Delta x_n \\ \Delta x_{n+1}\end{bmatrix} \\ & + \begin{bmatrix}\nabla_{x_n}\tilde{V}_n(\hat{x}_{n|m}, \hat{x}_{n+1|m}) \\ \nabla_{x_{n+1}}\tilde{V}_n(\hat{x}_{n|m}, \hat{x}_{n+1|m})\end{bmatrix}^\top \begin{bmatrix}\Delta x_n \\ \Delta x_{n+1}\end{bmatrix} \end{aligned} \tag{7}$$

where we used $\Delta x_n := x_n - \hat{x}_{n|m}$, $\Delta x_{n+1} := x_{n+1} - \hat{x}_{n+1|m}$ and $B_n(\hat{x}_{n|m}, \hat{x}_{n+1|m}) \approx \nabla^2 \tilde{V}_n(\hat{x}_{n|m}, \hat{x}_{n+1|m})$ is *some* approximation of the Hessian of $\tilde{V}_n$ at the current estimate $(\hat{x}_{n|m}, \hat{x}_{n+1|m})$. For the following derivations to be well-defined, the Hessian approximation has to be positive definite.

Assumption 1. The Hessian approximation $B_n(x_n, x_{n+1})$ is symmetric and positive definite for all x_n, x_{n+1}.

Note that $\nabla_{x_n}\tilde{V}_n(x_n, x_{n+1})$ is equivalent to the gradient of the objective function $V_n(x_n, \ldots, x_m)$ of the MHE problem w.r.t. x_n. We thus have $\nabla_{x_n}\tilde{V}_n(\hat{x}_{n|m}, \hat{x}_{n+1|m}) = 0$, since $\hat{x}_{n|m}, \hat{x}_{n+1|m}$ is a solution to the MHE problem. We can thus reformulate (7) as:

$$\tilde{V}_n^{\text{QUAD}}(x_n, x_{n+1}) = \frac{1}{2}\begin{bmatrix}1 \\ \Delta x_n \\ \Delta x_{n+1}\end{bmatrix}^\top \begin{bmatrix} c_n & 0 & q_n^\top \\ 0 & E_n & S_n^\top \\ q_n & S_n & D_n \end{bmatrix} \begin{bmatrix}1 \\ \Delta x_n \\ \Delta x_{n+1}\end{bmatrix}$$

where

$$c_n = 2\tilde{V}_n(\hat{x}_{n|m}, \hat{x}_{n+1|m}), \quad q_n = \nabla_{x_{n+1}}\tilde{V}_n(\hat{x}_{n|m}, \hat{x}_{n+1|m}),$$

and

$$B_n(\hat{x}_{n|m}, \hat{x}_{n+1|m}) =: \begin{bmatrix} E_n & S_n^\top \\ S_n & D_n \end{bmatrix}.$$

Applying the Schur complement lemma, we can reformulate $\tilde{V}_n^{\text{QUAD}}(x_n, x_{n+1})$ as

$$\begin{aligned} \tilde{V}_n^{\text{QUAD}}(x_n, x_{n+1}) &= q_n^\top \Delta x_{n+1} + \frac{1}{2}\Delta x_{n+1}^\top P_{n+1}^{-1} \Delta x_{n+1} \\ &+ \frac{1}{2}\|\Delta x_n + E_n^{-1} S_n^\top \Delta x_{n+1}\|^2_{E_n^{-1}} + \text{const} \end{aligned}$$

where

$$P_{n+1} := \left(D_n - S_n E_n^{-1} S_n^\top\right)^{-1}. \tag{8}$$

In this form, it is easy to minimize $\tilde{V}_n^{\text{QUAD}}(x_n, x_{n+1})$ wrt. x_n. The solution map is $x_n^*(x_{n+1}) = \hat{x}_{n|m}$ - $E_n^{-1} S_n^\top \Delta x_{n+1}$. We thus have

$$\begin{aligned} \alpha_{n+1}(x_{n+1}) &= \tilde{V}_n^{\text{QUAD}}(x_n^*(x_{n+1}), x_{n+1}) \\ &= \frac{1}{2}\|x_{n+1} - \breve{x}_{n+1}\|^2_{P_{n+1}^{-1}} + \text{const} \end{aligned}$$

with

$$\breve{x}_{n+1} = \hat{x}_{n+1|m} - P_{n+1}\nabla_w \psi(\hat{w}_{n|m}) \tag{9}$$

where $\hat{w}_{n|m} = \hat{x}_{n+1|m} - f_n(\hat{x}_{n|m})$.

Proposition 1. *If Assumption 1 is satisfied, the updated arrival cost covariance matrix P_{n+1} is positive definite.*

Proof. This follows directly from the fact that the inverse covariance matrix P_{n+1}^{-1} is given as the Schur complement of the block E_n in $B_n(\hat{x}_{n|m}, \hat{x}_{n+1|m})$. □

Theorem 2. *The arrival cost update defined in* (9)*,* (8) *satisfies the gradient condition for the arrival cost [8].*

Proof. Using (9), we can express the gradient of the updated arrival cost $\alpha_{n+1}(x_{n+1})$ as:

$$\begin{aligned} &\nabla \alpha_{n+1}(x_{n+1}) \\ &\quad = P_{n+1}^{-1}\left(x_{n+1} - \hat{x}_{n+1|m}\right) + \nabla\psi\left(\hat{x}_{n+1|m} - f_n(\hat{x}_{n|m})\right). \end{aligned}$$

Evaluating the gradient at the estimate $\hat{x}_{n+1|m}$, we obtain

$$\nabla\alpha_{n+1}(\hat{x}_{n+1|m}) = \nabla\psi\left(\hat{x}_{n+1|m} - f_n(\hat{x}_{n|m})\right), \quad (10)$$

which shows that the gradient condition is satisfied. □

Corollary 1. *We assume ψ and ϕ are convex. If a Generalized Gauss-Newton (GGN) Hessian approximation is used within the quadratic approximation in* (7)*, the arrival cost update in* (9)*,* (8) *is given explicitly as:*

$$P_{n+1} = \hat{Q}_n + \hat{A}_n F_n^{-1} \hat{A}_n^\top \quad (11)$$

$$F_n = P_n^{-1} + \hat{C}_n^\top \hat{R}_n^{-1} \hat{C}_n \quad (12)$$

$$\breve{x}_{n+1} = \hat{x}_{n|m} - P_{n+1}\nabla_w \psi(\hat{w}_{n|m}). \quad (13)$$

where $\hat{A}_n := \frac{\partial f_n}{\partial x_n}(\hat{x}_{n|m})$, $\hat{C}_n := \frac{\partial h_n}{\partial x_n}(\hat{x}_{n|m})$, *and*

$$\hat{w}_{n|m} := \hat{x}_{n+1|m} - f_n(\hat{x}_{n|m}), \quad \hat{Q}_n^{-1} := \nabla_w^2 \psi(\hat{w}_{n|m}),$$
$$\hat{v}_{n|m} := h(\hat{x}_{n|m}) - y_n, \quad \hat{R}_n^{-1} := \nabla_v^2 \phi(\hat{v}_{n|m}).$$

Proof. The GGN Hessian approximation of $\nabla^2 \tilde{V}_n(x_n, x_{n+1})$ is given as $B_n^{\text{GGN}}(x_n, x_{n+1}) = J_n(x_n)^\top \Lambda(v_n, w_n) J_n(x_n)$ with

$$\Lambda(v_n, w_n) = \begin{bmatrix} P_n^{-1} & & \\ & \nabla^2\phi(v_n) & \\ & & \nabla^2\psi(w_n) \end{bmatrix}, \quad (14)$$

$$J(x_n) = \begin{bmatrix} \mathbb{1} & \mathbb{0} \\ C_n(x_n) & \mathbb{0} \\ -A_n(x_n) & \mathbb{1} \end{bmatrix} \quad (15)$$

with $C_n(x_n) = \frac{\partial h_n}{\partial x_n}(x_n)$ and $A_n(x_n) = \frac{\partial f_n}{\partial x_n}(x_n)$. Note that the Jacobian $J(x_n)$ does not depend on x_{n+1}. Using this notation, the matrices E_n, D_n are given as

$$E_n = P_n^{-1} + \hat{A}_n^\top \hat{Q}_n^{-1} \hat{A}_n + \hat{C}_n^\top \hat{R}_n^{-1} \hat{C}_n, \qquad D_n = \hat{Q}_n^{-1},$$

and equation (8) is explicitly given as

$$P_{n+1}^{-1} = \hat{Q}_n^{-1} - \hat{Q}_n^{-1} \hat{A}_n^\top E_n^{-1} \hat{A}_n \hat{Q}_n^{-1}. \quad (16)$$

Applying the Woodbury identity to (16), we obtain the formula in (11). □

Please note that the arrival cost update in Corollary 1 corresponds to a Kalman Filter predict and update step, where the update step is done in information form.

Corollary 2. *Suppose the GGN Hessian is used as Hessian approximation $B_n(\hat{x}_{n|m}, \hat{x}_{n+1|m})$ within the arrival cost update. If P_n is positive definite and $\nabla^2\psi(w)$, $\nabla^2\psi(v)$ are positive definite for all w, v, then P_{n+1} is positive definite.*

Proof. From (15), it is easy to see that the Jacobian $J(x_n)$ has full rank for all x_n. Under the given assumptions, the matrix $\Lambda_n(w_n, v_n)$ is positive definite for all w_n, v_n, which implies that the GGN Hessian $B_n^{\text{GGN}}(x_n, x_{n+1})$ is positive definite as well. Proposition 1 then directly implies positive definiteness of P_{n+1}. □

Remark 1 (Nonconvex Loss Functions)**.** Outliers within the measured outputs can be effectively rejected by using a nonconvex loss function for the output error. One particular choice of a nonconvex loss function for outlier rejection is the following:

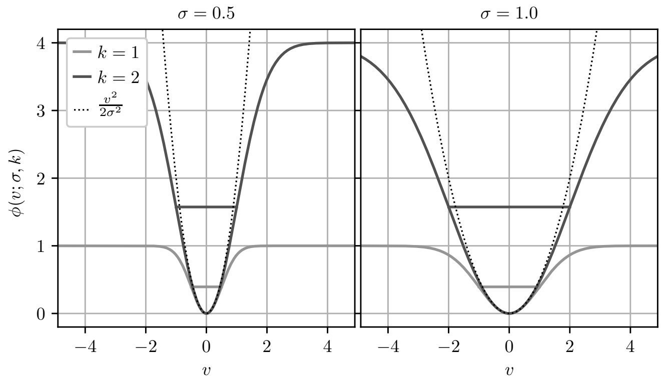

Fig. 1. Loss function $\phi(v;\sigma,k)$ for $\sigma = 0.5$ (left) and $\sigma = 1$ (right). For comparison the quadratic loss that matches the curvature at the origin is shown. The horizontal lines indicate the convex regions of the loss functions.

$$\phi(v;\sigma,k) = k^2\left(1 - \exp\left(-\frac{v^2}{2k^2\sigma^2}\right)\right).$$

The Hessian of $\phi(v;\sigma,k)$ evaluated at zero is given by $\phi''(0;\sigma,k) = \sigma^{-2}$, such that we can easily tune the above loss to approximate an L_2-loss in a neighborhood of the origin (cf. Fig. 1). The parameter k defines the interval on which the loss function is convex. In particular, $\phi(v;\sigma,k)$ is convex on $[-k\sigma, k\sigma]$.

In order for the arrival cost covariance matrix to be positive definite, we need to replace the Hessian $\hat{R}^{-1}$, within the arrival cost update by a positive definite approximation. To this end, we use an Extended Gauss-Newton (XGN) Hessian approximation as introduced in [12]. The XGN Hessian can be used both for convex and nonconvex loss functions that have a unique global minimum. The XGN Hessian leads to a quadratic approximation of the loss function that preserves the location of the global minimum at zero. We refer to [12] for a more elaborate analysis of the XGN approximation.

IV. Arrival Cost Regularization and Forgetting

In the following, we illustrate how the arrival cost update can be regularized in order to guarantee that the arrival cost covariance will be within a given lower and upper bound. Regularization of the arrival cost is particularly beneficial if the system reaches a steady state at which the (linearized) system is no longer detectable. In this case, the covariance associated with the state estimate might increase unboundedly leading to a very high sensitivity of the state estimate to outliers.

Let us first consider detectability of the augmented system, as given in (2), at a steady state $(x_{\text{ss}}, u_{\text{ss}})$. The Jacobian of the system dynamics and the output function at the steady state are given as

$$A_{\text{ss}} = \frac{\partial f}{\partial x}(x_{\text{ss}}, u_{\text{ss}}) = \begin{bmatrix} A_{\text{ss}}^z & A_{\text{ss}}^\theta \\ \mathbb{0} & \mathbb{1} \end{bmatrix}, \quad C_{\text{ss}} = \frac{\partial h}{\partial x}(x_{\text{ss}}) = \begin{bmatrix} C_{\text{ss}}^z & \mathbb{0} \end{bmatrix}.$$

Proposition 3 (Non-Detectability at the Steady State)**.** *We consider the linear system obtained by linearizing* (2) *at the steady state* $(x_{\text{ss}}, u_{\text{ss}})$*. This linear system is not detectable if*

(a) *rank* $(C^z_{\text{ss}}) < n_\theta$*, or*
(b) *rank* $\left(A^\theta_{\text{ss}}\right) < n_\theta$*.*

Proof. From Hautus Lemma [13], we know that the system is detectable if

$$\operatorname{rank}\left(\begin{bmatrix}\lambda\mathbb{1} - A_{\text{ss}} \\ C_{\text{ss}}\end{bmatrix}\right) = n_x \tag{17}$$

for all $\lambda \in \operatorname{eig}(A_{\text{ss}})$, $|\lambda| \geq 1$. The eigenvalues of A_{ss} are given as the solution of

$$\det\left(\begin{bmatrix}A^z_{\text{ss}} - \lambda\mathbb{1} & A^\theta_{\text{ss}} \\ \mathbb{0} & \mathbb{1} - \lambda\mathbb{1}\end{bmatrix}\right) = 0.$$

Since

$$\det\left(\begin{bmatrix}A^z_{\text{ss}} - \lambda\mathbb{1} & A^\theta_{\text{ss}} \\ \mathbb{0} & \mathbb{1} - \lambda\mathbb{1}\end{bmatrix}\right) = \det\left(A^z_{\text{ss}} - \lambda\mathbb{1}\right)\det\left(\mathbb{1} - \lambda\mathbb{1}\right),$$

we have $\operatorname{eig}(A_{\text{ss}}) = \operatorname{eig}(A^z_{\text{ss}}) \cup \operatorname{eig}(\mathbb{1})$. For our particular type of system, we thus need to check the rank of the matrix

$$\mathcal{M}(\lambda) := \begin{bmatrix}A^z_{\text{ss}} - \lambda\mathbb{1} & A^\theta_{\text{ss}} \\ \mathbb{0} & \mathbb{1} - \lambda\mathbb{1} \\ C^z_{\text{ss}} & \mathbb{0}\end{bmatrix} \tag{18}$$

for all $\lambda \in \operatorname{eig}(A_{\text{ss}}) = \operatorname{eig}(A^z_{\text{ss}}) \cup \operatorname{eig}(\mathbb{1})$ with $|\lambda| \geq 1$. Considering $\lambda = 1 \in \operatorname{eig}(\mathbb{1})$, we obtain

$$\operatorname{rank}\left(\mathcal{M}(1)\right) \leq \operatorname{rank}\left(C^z_{\text{ss}}\right) + \operatorname{rank}\left(\begin{bmatrix}A^z_{\text{ss}} - \lambda\mathbb{1} & A^\theta_{\text{ss}}\end{bmatrix}\right), \tag{19}$$

i.e. $\operatorname{rank}\left(\mathcal{M}(1)\right) \leq \operatorname{rank}\left(C^z_{\text{ss}}\right) + n_z < n_\theta + n_z = n_x$ if (a) holds. Similarly, we have

$$\operatorname{rank}\left(\mathcal{M}(1)\right) \leq \operatorname{rank}\left(\begin{bmatrix}A^z_{\text{ss}} - \mathbb{1} \\ C^z_{\text{ss}}\end{bmatrix}\right) + \operatorname{rank}\left(A^\theta_{\text{ss}}\right), \tag{20}$$

which implies $\operatorname{rank}\left(\mathcal{M}(1)\right) \leq n_z + \operatorname{rank}\left(A^\theta_{\text{ss}}\right) < n_z + n_\theta$ if (b) holds. □

Note that the above proposition directly implies that any augmented system is not detectable at the steady state if the number of measurements n_y is less than the number of parameters n_θ.

At a steady state where some of the states or parameters are not detectable, the arrival cost covariance P_n will grow with each update step due to the state noise covariance. In the following, we propose one way of introducing regularization to the arrival cost update to prevent the covariance from growing unboundedly.

Definition 1 (Regularization of the Arrival Cost Covariance)**.** *A regularized arrival cost covariance matrix,* $\bar{P}_{n+1}$*, is obtained with the following formulas:*

$$\bar{P}_{n+1} = \hat{Q}_n + \bar{Q} + \hat{A}_n\,\bar{F}_n^{-1}\hat{A}_n^\top \tag{21}$$

$$\bar{F}_n = \bar{P}_n^{-1} + \hat{C}_n^\top\hat{R}_n^{-1}\hat{C}_n + \bar{C}^\top\bar{R}^{-1}\bar{C} \tag{22}$$

for fixed matrices $\bar{Q} \in \mathbb{R}^{n_x \times n_x}$*,* $\bar{R} \in \mathbb{R}^{\bar{n}_y \times \bar{n}_y}$*,* $\bar{C} \in \mathbb{R}^{\bar{n}_y \times n_x}$ *and with* $\bar{R}$ *positive definite and* $\bar{Q}$ *positive semidefinite.*

The regularized arrival cost update can be interpreted as a GGN-based arrival cost update with an additional pseudo-measurement $\bar{y}_n = \bar{C}\hat{x}_{n|m}$. We would typically add pseudo-measurements of the estimated parameters θ_n, i.e. choose $\bar{C} = [\mathbb{0}_{n_\theta \times n_z}\ \mathbb{1}_{n_\theta \times n_\theta}]$.

Proposition 4. *The regularized arrival cost update defined in* (21) *and* (22) *can be obtained as the solution of the following variant of problem* (6)*.*

$$\alpha_{n+1}(x_{n+1}) = \min_{x_n}\ \tilde{V}_n^{\text{QUAD}}(x_n, x_{n+1}) + \frac{1}{2}\|\bar{C}x_n - \bar{y}_n\|^2_{\bar{R}^{-1}}$$

where $\bar{y}_n = \bar{C}\hat{x}_{n|m}$ *and* $\tilde{V}_n^{\text{QUAD}}$ *is given by a quadratic approximation of* $\tilde{V}_n$ *with Hessian* $\bar{B}_n^{\text{GGN}}$ *given as*

$$\bar{B}_n^{\text{GGN}}(x_n, x_{n+1}) = J_n(x_n)^\top\nabla_s^2\bar{\Lambda}_n\left(r_n(x_n, x_{n+1})\right)J_n(x_n)$$

with $\nabla_s^2\bar{\Lambda}_n(s) = \operatorname{diag}\left(P_n^{-1}, \hat{R}_n^{-1}, (\hat{Q}_n + \bar{Q})^{-1}\right)$*.*

Proof. Note that the gradient of the term $\frac{1}{2}\|\bar{C}x_n - \bar{y}_n\|^2_{\bar{R}^{-1}}$ with respect to x_n is zero and thus only the Hessian of the regularized arrival cost is changed. In particular, the Hessian w.r.t. x_n, denoted by E_n, is given by

$$E_n = P_n^{-1} + \hat{A}_n^\top(\hat{Q}_n + \bar{Q})^{-1}\hat{A}_n + \hat{C}_n^\top\hat{R}_n^{-1}\hat{C}_n + \bar{C}^\top\bar{R}^{-1}\bar{C}.$$

□

Proposition 5. *For a positive definite matrix* $\bar{Q}$ *and a positive definite matrix* $\bar{R}$*, the regularization of the arrival cost implies the following bound:*

$$\bar{C}^\top\bar{R}^{-1}\bar{C} \preceq \bar{F}_n \preceq \bar{Q}^{-1} + \hat{C}_n^\top\hat{R}_n^{-1}\hat{C}_n + \bar{C}^\top\bar{R}^{-1}\bar{C}.$$

Proof. With $\bar{P}_n^{-1}$ and $\hat{C}_n^\top\hat{R}^{-1}\hat{C}_n$ positive semidefinite, the lower bound follows from (22). Noting that $\bar{Q} \preceq \bar{P}_n$ and thus $\bar{Q}^{-1} \succ \bar{P}_n^{-1}$, we obtain the upper bound from (22). □

Note that the introduction of the matrix $\bar{Q}$ can be interpreted as a form of *forgetting*, as we have $\bar{P}_{n+1} \succeq \bar{Q}$. Increasing $\bar{Q}$ leads to a larger arrival cost covariance and thus increased *forgetting*. An alternative way of introducing forgetting within the arrival cost update would be to downweight the inverse covariance matrix directly, i.e. using $\gamma\bar{P}_{n+1}^{-1}$ with $0 < \gamma < 1$.

In particular for estimation problems where both the system state and additional system parameters are estimated, introducing forgetting is advantageous if the parameters can be assumed to stay constant for most of the time, but might be subject to step changes. On the other hand, forgetting increases the sensitivity to outliers if not enough information is available within the measurements of the current optimization window to accurately estimate the parameter value.

Based on these considerations, we propose to use an adaptive regularization scheme for the arrival cost update when estimating both the system state and parameters. The main idea is to introduce pseudo-measurements of the estimated parameters within the arrival cost update only if the variance associated with these estimates does not decrease sufficiently within the MHE optimization window. To this end, let us first consider an upper bound on the growth of the covariance associated with the parameter estimates within the optimization window.

Lemma 6. *We consider an augmented system for combined state and parameter estimation as in* (2)*. Suppose* $(\hat{x}_{n|m}, \ldots, \hat{x}_{m|m})$ *is the solution to the current MHE problem with arrival cost defined by* $\breve{x}_n$*,* P_n*.*

Let $\hat{Q}_i = \mathrm{diag}(\hat{Q}_i^z, \hat{Q}_i^\theta) = \nabla_w^2 \psi(\hat{w}_{i|m})^{-1}$*,* $i = n, \ldots, m$*. We consider the covariance matrix* $P_{m+1|m}$ *associated with the prediction* $\hat{x}_{m+1|m} = f_m(\hat{x}_{m|m})$*. We have*

$$P^\theta_{m+1|m} \preceq P^\theta_n + \sum_{i=n}^{m} \hat{Q}^\theta_n.$$

where $P^\theta_{m+1|m}$ *and* P^θ_n *are the diagonal blocks of* $P_{m+1|m}$ *and* P_n *that correspond to the etimated parameters* θ*.*

Proof. We obtain $P_{m+1|m}$ by applying the recursion in (11), (12) N-times starting from $P_{n|m} = P_n$ and using the linearization of the MHE problem at the solution [14]–[16].

Let us consider the covariance $P_{n+1} = P_{n+1|m}$ after applying the recursion once. From (12), it directly follows that $F_n \succeq P_n^{-1}$. We thus have

$$P_{n+1} \preceq \hat{Q}_n + \hat{A}_n P_n \hat{A}_n^\top$$

Due to the particular structure of $\hat{A}_n$ resulting from (2), we obtain

$$P_{n+1} = \begin{bmatrix} P^z_{n+1} & P^{\theta,z}_{n+1} \\ (P^{\theta,z}_{n+1})^\top & P^\theta_{n+1} \end{bmatrix} = \begin{bmatrix} * & * \\ * & \hat{Q}^\theta_n + P^\theta_n \end{bmatrix}.$$

□

Taking this upper bound as a reference value for the growth of the covariance associated with the estimated parameters within the optimization window for the case when no measurements are available, we can define an adaptive arrival cost regularization as outlined next.

Definition 2 (Adaptive Arrival Cost Regularization)**.** *We consider an augmented system for combined state and parameter estimation as in* (2)*. Let* $\bar{C} = [\mathbb{0}_{n_\theta \times n_z} \; \mathbb{1}_{n_\theta \times n_\theta}]$ *and suppose* $\tilde{R} = \mathrm{diag}(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_{n_\theta}^2)$ *is positive definite. We furthermore assume* $\hat{Q}_i = \mathrm{diag}(\hat{Q}_i^z, \hat{Q}_i^\theta) := \nabla_w^2 \psi(\hat{w}_{i|m})^{-1}$*,* $i = n, \ldots, m$*.*

We set the pseudo-measurement covariance in (22) *to*

$$\bar{R}_n = \mathrm{diag}\left(\frac{\tilde{\sigma}_1^2}{\kappa_1}, \ldots, \frac{\tilde{\sigma}_{n_\theta}^2}{\kappa_{n_\theta}}\right) \tag{23}$$

with

$$\kappa_j = \frac{\hat{\sigma}^2_{m+1|m,j}}{\hat{\sigma}^2_{n|m,j} + \sum_{i=n}^{m} Q^\theta_{i;j,j}} \tag{24}$$

where $\hat{Q}^\theta_{i;j,j}$ *denotes the* j*-th diagonal entry of the matrix* $\hat{Q}^\theta_i$*,* $\hat{\sigma}^2_{n|m,j}$ *and* $\hat{\sigma}^2_{m+1|m,j}$ *denote the* j*-th diagonal entry of* $P_n = P_{n|m}$ *and* $P_{m+1|m}$ *respectively.*

Note that $0 \leq \kappa_j \leq 1$ due to Lemma 6 and thus the pseudo measurements have minimal covariance $\tilde{\sigma}_j^2$ if no information about the parameters is available within the optimization window, i.e. we have $\hat{\sigma}^2_{m+1|m,j} = \hat{\sigma}^2_{n|m,j} + \sum_{i=n}^{m} Q^\theta_{i;j,j}$ and

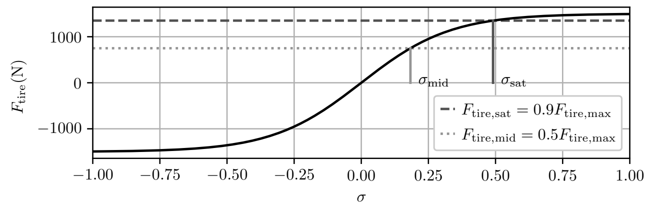

Fig. 2. Nonlinear dependency of the tire slip σ and the transmitted force F_{tire}. The values σ_{mid} and σ_{sat} are related to the slip values of either 50% or 90% of the maximum force that can be transmitted.

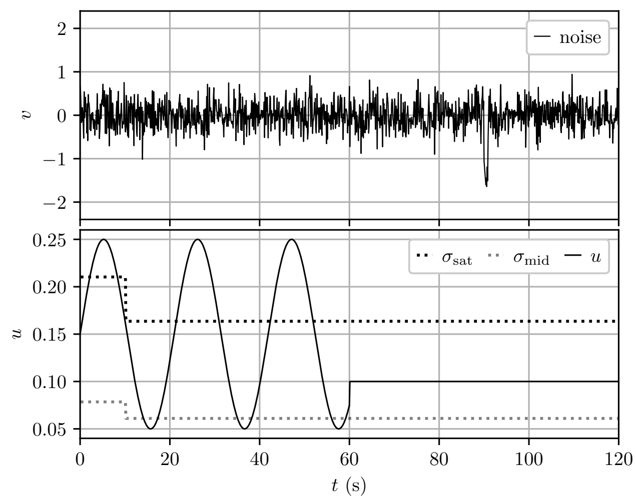

Fig. 3. Open-loop simulation: The plot in the top row show the measurement noise v with a series of outlier at starting at $t = 85$s. The bottom plot shows the open loop control input u.

thus $\kappa_j = 1$. If however $\kappa_j \ll 1$, the covariance of the pseudo-measurements tends to infinity and no regularization is added to the updated covariance.

V. Application: Longitudinal Vehicle Motion Estimation

We demonstrate the proposed methods on a longitudinal vehicle motion and tire friction parameter estimation example. To this end, we consider an open-loop scenario to show the estimation improvement isolated from feedback actions.

A. Longitudinal Vehicle Motion Model

We use a longitudinal vehicle motion model where we measure the longitudinal position p and our aim is to control the speed s in a steady-state. The system is characterized by the state $z = [p \quad s]$. As a single input $u = [\sigma]$, the model takes the tire slip-ratio σ

$$\sigma = \frac{r\omega - s}{s}, \tag{25}$$

which is the ratio of how much the tire velocity at the contact point $r\omega$ deviates from the ground velocity, where r is the effective tire radius and ω is the rotational speed. We model the tire transmission force F_{tire} by means of a parameterized

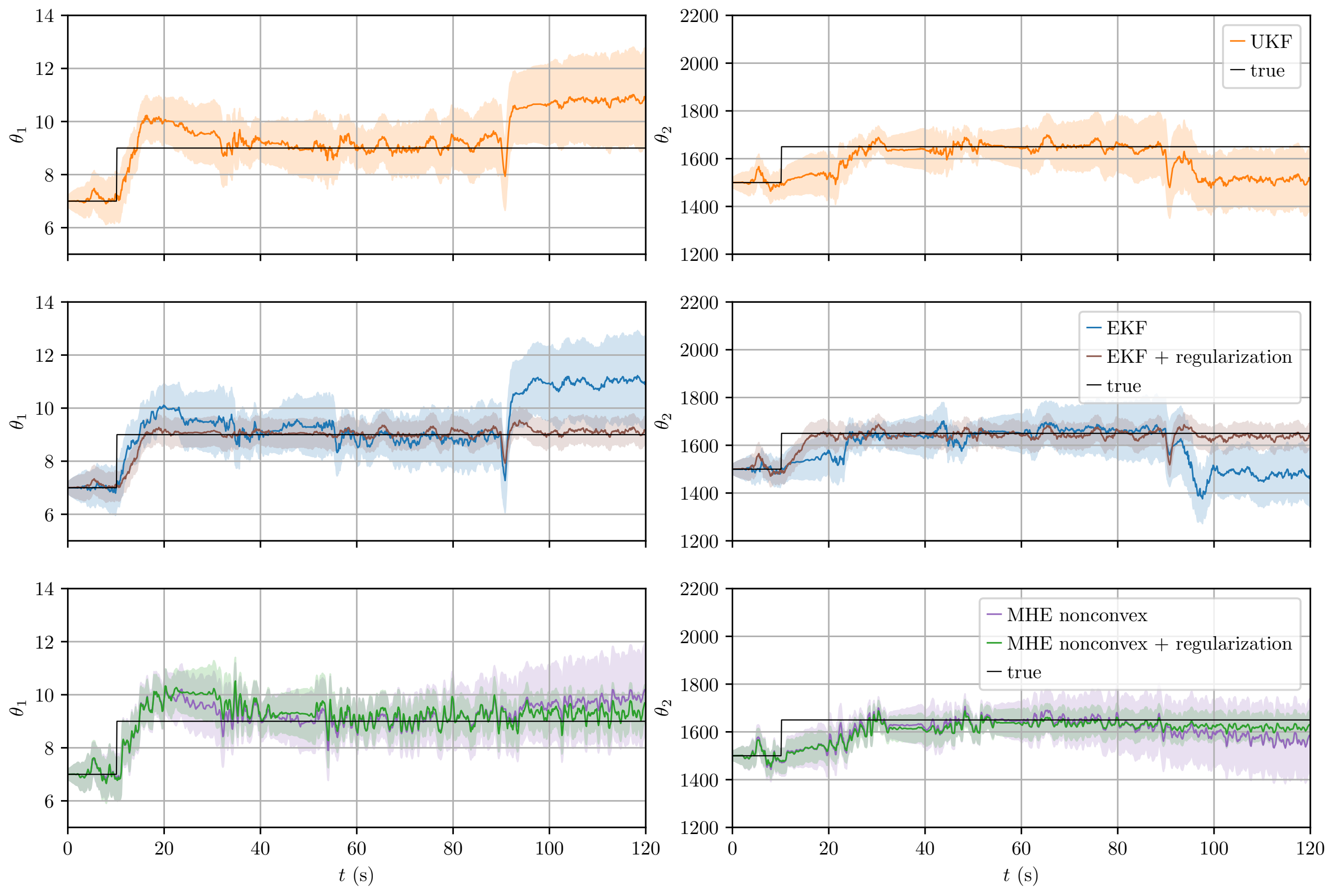

Fig. 4. Parameter estimates obtained in an offline estimation problem for the noise and control trajectory shown in Fig. 3. Results obtained with an UKF (top), EKF with and without regularization (middle) and MHE with and without regularization (bottom) are shown. The uncertainty band indicate a $\pm 1\sigma$ interval associated with the respective estimates. The assumed state and measurement noise covariances are the same for the EKF, UKF and the arrival cost updates. Additional pseudo-measurements are used for the methods using regularization.

function with two unknown and changing parameters $\theta = [\theta_1 \quad \theta_2]$ given by

$$F_{\text{tire}}(\sigma) = \theta_2 \tanh(\theta_1 \sigma). \tag{26}$$

The choice of this model is motivated in [11]. The shape of the tire transmission force function simplifies the estimation problem compared to more elaborate functions like the *Magic Formula* [17] with a potential drawback constituting in the fact that it can not capture the difference between dry and lubricate friction. Nevertheless, the decreasing part of the force transmission curve related to the insufficiently modeled part is not within the operating region of a usual vehicle control algorithm and the used model is sufficient to present the effectiveness of the proposed algorithms. Fig. 2 shows the tire force transmission function. Input values σ_{mid} and σ_{sat} relate to slip values that either transmit half or 90% of the maximum achievable transmission force. The parameters of the rolling resistance force $F_{\text{roll}} = \text{sign}(s)\bar{F}_{\text{roll}}$ and the nonlinear air drag force $F_{\text{air}} = c_{\text{air}} s^2$ are assumed to be known. The motion of the model can be described by

$$m\ddot{p} = F_{\text{tire}}(\sigma) - F_{\text{roll}}(s) - F_{\text{air}}(s), \tag{27}$$

which we reformulate and simplify for positive velocities into the first order differential equations

$$\dot{z} = f_{\text{c}}(z, u, \theta) = \begin{bmatrix} s \\ \frac{1}{m}(\theta_2 \tanh(\theta_1 \sigma) - \bar{F}_{\text{roll}} - c_{\text{air}} s^2) \end{bmatrix}. \tag{28}$$

An RK4 integrator is used to integrate (28) to obtain the discrete state mapping $z_{k+1} = \tilde{f}(z_k, u_k, \theta)$. We use the augmented model as presented in Section II in order to include the system parameters θ_2 and θ_1 within the discrete system dynamics. With measuring the position p solely, we have $h(x_k) = p_k$.

B. Methods and Tuning Parameters

In addition to several variants of the MHE method, we consider an Extended Kalman Filter (EKF) and an Unscented Kalman Filter (UKF). Both a standard EKF and an EKF with additional pseudo-measurements is used. For all methods, we assume a measurement variance of $R = 0.1$, which corresponds to the true measurement noise variance in the simulations. For the EKF, UKF and the arrival cost update, the state noise is assumed to have a covariance

$$\bar{Q} = \text{diag}(10^{-5}, 5 \cdot 10^{-5}, 0.008, 60).$$

For MHE, we use an L_2 loss for the state noise with $Q = \mathrm{diag}(10^{-5}, 5 \cdot 10^{-5}, 0.004, 30)$, i.e. the variance is reduced for the parameters. For the measurement noise, we use a negative Gaussian loss with $\sigma = R = 0.1$ and $k = 1$ (cf. Fig. 1). The MHE horizon is set to $N = 6$.

For all variants using regularization via pseudo-measurements, we set the covariance of the pseudo measurements to $\bar{R} = \mathrm{diag}(0.08, 0.000005)$ and $\bar{C} = [\mathbb{0}_2 \ \mathbb{1}_2]$, i.e. we add pseudo-measurements for the parameters only.

C. Simulation Results

In an open-loop simulation, we compare how the presented approaches perform on the longitudinal vehicle model in the presence of outliers. The input trajectory as well as the measurement noise is shown in Fig. 3. The system is excited in the first part of the simulation until $t = 60$s and then kept at a steady state. The excitation is performed in the range of the major non-linearity of the system which is indicated by means of the dotted lines related to significant points in the tanh-function. The top plot in Fig. 3 shows the measurement noise trajectory, including a series of outliers starting at time $t = 85$s. The estimated parameters for this scenario obtained with the different methods are shown in Fig. 4. Both standard variants (EKF, UKF) get severely disturbed by the outliers, whereas the MHE variants are not affected as severly. Furthermore, the MHE variant with regularization leads to bounded covariance estimates, even though no system excitation is present after $t = 60s$, except for the measurement noise. For the MHE variant without regularization, the covariance keeps growing and the estimates start to drift as is the case for the UKF and EKF.

The true speed s of the system approaches a steady state value of $s_{\mathrm{ss}} = 27.87 \frac{\mathrm{m}}{\mathrm{s}}$ at the end of the simulation. Due to the linear output function, the observability matrix $\mathcal{O}$ is independent of the position p (which does not reach a steady state). The singular values of the observability matrix at the steady state are given by

$$\sigma(\mathcal{O}) = \{2.02,\ 2.19 \cdot 10^{-1},\ 8.08 \cdot 10^{-4},\ 2.25 \cdot 10^{-22}\},$$

which clearly indicates that the system is not observable.

VI. Conclusions and Outlook

We considered a Moving Horizon Estimation (MHE) approach for combined state and parameter estimation in the presence of outlier measurements. In this setting, nonconvex loss functions are of particular interest as they can be used to effectively reject outlier measurements in an automated, model-based way. We thus proposed an arrival cost update formula for general convex and some nonconvex loss functions that is based on a generalized Gauss-Newton (GGN) or an extended Gauss-Newton (XGN) Hessian approximation. If additional system parameters are estimated alongside the state, the estimation problem can become ill-posed if the system reaches a steady state as the parameters might then be no longer observable. To this end, we introduced an adaptive regularization scheme that comprises both forgetting and additional pseudo-measurements of the estimated parameters. Using a simplified longitudinal vehicle motion estimation problem, we compared the proposed MHE algorithm to the extended and unscented Kalman Filter and showed superior estimation accuracy.

Currently, we assume that the MHE problem is solved exactly for deriving the arrival cost update. Future research will consider a real-time variant of the MHE algorithm where only an approximate solution of the MHE problem is available. Furthermore, future work will include an efficient implementation of the proposed algorithm as well as extensive benchmarking.

References

[1] C. Rao and J. Rawlings, "Nonlinear Moving Horizon State Estimation," in *Nonlinear Predictive Control*, F. Allgöwer and A. Zheng, Eds. Basel Boston Berlin: Birkhäuser, 2000, pp. 45–69.

[2] C. Rao, J. Rawlings, and J. Lee, "Constrained linear state estimation - a moving horizon approach," *Automatica*, vol. 37, no. 2, pp. 1619–1628, 2001.

[3] M. Tenny and J. Rawlings, "Efficient Moving Horizon Estimation and Nonlinear Model Predictive Control," in *Proceedings of the American Control Conference*, Anchorage, AK, 2002.

[4] 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.

[5] M. Diehl, H. J. Ferreau, and N. Haverbeke, "Efficient numerical methods for nonlinear MPC and moving horizon estimation," in *Nonlinear model predictive control*, ser. Lecture Notes in Control and Information Sciences, L. Magni, M. Raimondo, and F. Allgöwer, Eds. Springer, 2009, vol. 384, pp. 391–417.

[6] R. Dutter, *Robust regression: Different approaches to numerical solution and algorithms*. Fachgruppe für Statistik, Eidgenössische Techn. Hochsch., 1975.

[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] K. Baumgärtner, A. Zanelli, and M. Diehl, "A gradient condition for the arrival cost in moving horizon estimation," in *Proceedings of the European Control Conference (ECC) 2020*, 2020.

[9] M. Zanon, J. V. Frasch, and M. Diehl, "Nonlinear moving horizon estimation for combined state and friction coefficient estimation in autonomous driving," in *Proceedings of the European Control Conference (ECC)*, 2013, pp. 4130–4135.

[10] H. B. Pacejka, *Tyre and Vehicle Dynamics*. Elsevier, 2006.

[11] O. Jonson and E. Enders, "Correlation work on shaker rig tests and simulations – an investigation of damper, bushing, friction and tire modeling with respect to vertical vibration insulation," Ph.D. dissertation, 07 2016.

[12] K. Baumgärtner and M. Diehl, "The extended Gauss-Newton method for nonconvex loss functions and its application to time-optimal model predictive control," in *Proceedings of the American Control Conference (ACC)*, 2022.

[13] M. Hautus, "Controllability and stabilizability of sampled systems," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 528–531, 1972.

[14] R. López-Negrete and L. T. Biegler, "A moving horizon estimator for processes with multi-rate measurements: A nonlinear programming sensitivity approach," *Journal of Process Control*, vol. 22, no. 4, pp. 677–688, 2012.

[15] A. Bansal, M. Bhushan, and L. T. Biegler, "Covariance computation in MHE: A nonlinear regression approach," in *12th IEEE International Conference on Control and Automation (ICCA)*, 2016, pp. 663–668.

[16] K. Baumgärtner, J. Frey, R. Hashemi, and M. Diehl, "Zero-order moving horizon estimation for large-scale nonlinear processes," *Computers & Chemical Engineering*, vol. 154, p. 107433, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098135421002118

[17] H. B. Pacejka and E. Bakker, "The magic formula tyre model," *Vehicle System Dynamics*, vol. 21, no. sup001, pp. 1–18, 1992. [Online]. Available: https://doi.org/10.1080/00423119208969994