A proximity moving horizon estimator for a class of nonlinear systems

Meriem Gharbi | Christian Ebenbauer

Institute for Systems Theory and Automatic Control, University of Stuttgart, Stuttgart, Germany

Correspondence
Meriem Gharbi, Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, 70550 Stuttgart, Germany.
Email: meriem.gharbi@ist.uni-stuttgart.de

Funding information
Deutsche Forschungsgemeinschaft, Grant EB 425/6-1

Summary
In this article, we present a proximity-based formulation for moving horizon estimation (MHE) of a class of constrained discrete-time nonlinear systems. The cost function of the proposed estimator includes a convex stage cost as well as a Bregman distance for the a priori estimate. This rather general formulation of the cost function allows for a flexible design of the estimator depending on the setting of the problem at hand. We derive sufficient conditions for the global exponential stability of the underlying estimation error. We also investigate the robustness properties of the proposed MHE scheme applied for linear systems subject to unknown process and measurement disturbances in terms of an input-to-state stability analysis. The estimator performance is illustrated by means of simulation examples.

KEYWORDS
input-to-state stability, moving horizon estimation, proximal algorithms

1 | INTRODUCTION

Moving horizon estimation (MHE) is a powerful online optimization-based approach that estimates the state of a dynamical system by solving at each time instant a suitable optimization problem. Hereby, MHE takes a fixed and limited number of recent measurements into account and the considered horizon of measurements is shifted forward in time (in a receding horizon manner) as a new measurement becomes available. The main advantages of MHE include the ability to specify a performance criterion that can be designed specifically for the problem at hand, the ability to handle potential constraints on the estimated states in the underlying optimization problem, and a tractable online computational complexity compared to the full information estimation problem. The stability properties of the estimator depend on how to incorporate information into old measurements that are discarded by the receding-horizon optimization problem. Several design methods were proposed in the literature in order to preserve stability for linear systems and for nonlinear systems as well as for large-scale systems. We also mention works where robust stability of MHE is analyzed for linear and nonlinear systems, as well as methods for the fast implementation of MHE with stability guarantees. Typically, a quadratic prior weighting centered at a successively updated a priori estimate is introduced in the cost function in order to ensure stability. One particularly appealing strategy relevant to our approach solves a least-squares estimation problem where a stabilizing preestimating observer is employed into the a priori estimate and forward prediction equations, which allows ensuring design stability of the estimation error.
In our previous work,\textsuperscript{17} we introduced for the problem class of discrete-time linear systems a novel \textit{proximal} MHE formulation whose stability is inherited from the a priori estimate and in which a rather general and not necessarily quadratic performance criterion enables to handle outliers and to consider other specific data characteristics. More specifically, the cost function of the underlying optimization problem consists of general convex stage costs that enable to include, for example, sparsity promoting $\ell_1$ estimates, which have been highly successful in many signal processing applications.\textsuperscript{18,19} Furthermore, a quadratic prior weighting is included as a proximity measure between the MHE estimate and the a priori estimate chosen as a stabilizing Luenberger observer. Recently, we generalized this MHE formulation by introducing the so-called Bregman distance as a more general proximity measure for the stabilizing a priori estimate.\textsuperscript{20} Within this Bregman formulation, it is possible to transform a constrained estimation problem with a specific choice of the Bregman distance, in particular with the so-called relaxed barrier functions, into an unconstrained problem, which can, therefore, yield to a simple and efficient numerical implementation.\textsuperscript{21,22}

The goal of this work is to generalize these recent results.\textsuperscript{17,20} In particular, the contributions of this article and the main differences with respect to our recent results are twofold. First, we extend these results by investigating the inherent robustness properties of the proximity-based MHE scheme in terms of an input-to-state stability analysis. Second, we extend them to nonlinear systems. More specifically, we consider a class of nonlinear discrete-time systems that can be transformed into systems that are affine in the unmeasured state. We design a proximal MHE scheme based on Bregman distances for the resulting class of systems and derive sufficient conditions for the global exponential stability of the estimation error based on a particular assumption on the Bregman distance. Second, we investigate for a special case the input-to-state stability properties of the employed estimator, in which the Bregman distance is chosen as a weighted quadratic function. The main methodological concepts for the design and analysis within the framework of the proposed proximity-based MHE approaches are from the field of state estimation and observability theory and the powerful class of proximal point algorithms\textsuperscript{23-25} and proximal methods based on Bregman distances.\textsuperscript{26,27}

The remainder of this article is structured as follows. In Section 2, we introduce the considered class of nonlinear systems as well as the proposed MHE scheme. In Section 3, we analyze the nominal stability properties of the underlying estimation error. In Section 4, we restrict our attention to a special case of the considered class of systems for which the input-to-state stability analysis of the MHE strategy is investigated and established under rather mild assumptions. In Section 5, we illustrate our results by means of simulation examples.

\textbf{Notation:} Let $\mathbb{R}_+$ and $\mathbb{R}_{++}$ be the sets of nonnegative real and positive real numbers, respectively, and $S^n_+$ and $S^n_{++}$ be the sets of symmetric positive-semidefinite and positive-definite matrices of dimension $n \in \mathbb{N}_+$, respectively. For a vector $v \in \mathbb{R}^n$, let $\|v\|_p := \sqrt[n]{v^T P v}$ for any $P \in S^n_+$. Let $\|M\| = \sqrt{\lambda_{\max}(M^T M)}$ refer to the spectral norm of a matrix $M$. For a sequence of vectors $v = \{v_i, v_{i+1}, \ldots\}$, we define the norm $\|v_{i:k}\| := \sup_{S \subseteq \{i, i+1, \ldots, k\}} \{\|v_k\|\}$ for given $i, k \in \mathbb{N}$, as well as the norm of the column vector $\|v\| := \|\text{col}(v_i, v_{i+1}, \ldots, v_k)\|$. Moreover, let $\prod_{i=1}^k A_i := A_k \ldots A_2 A_1$ denote the left multiplication of the matrices $A_i$ for $i = 1, \ldots, k$.

\section{Formulation of the Estimator}

We consider a class of discrete-time nonlinear systems of the form

\begin{align}
\tag{1a}
\z_{k+1} &= f(\z_k, u_k), \\
\tag{1b}
y_k &= h(\z_k),
\end{align}

where $\z_k \in \mathbb{R}^n$ denotes the state vector, $u_k \in \mathbb{R}^m$ is the input vector, and $y_k \in \mathbb{R}^p$ is the output vector. We assume that (1) can be transformed into the following system:

\begin{align}
\tag{2a}
x_{k+1} &= A(y_k) x_k + \phi(y_k) + B(y_k) u_k, \\
\tag{2b}
y_k &= C x_k,
\end{align}

where $\phi$ denotes a nonlinearity that depends on the output $y_k$. We state this more formally in the following assumption.
Assumption 1. We assume that system (1) is globally equivalent to a linear observable system (2) via a diffeomorphism $x = T(z)$.

Necessary and sufficient conditions for Assumption 1 to hold can be found in Reference 28. We also assume that the state $x_k$ of the new system lies in the interior of the following polytopic set

$$x_k \in \mathcal{X} := \{ x \in \mathbb{R}^n : C_x x \leq d_x \},$$

where $C_x \in \mathbb{R}^{q \times n}$ and $d_x \in \mathbb{R}^q$, with $q_x \in \mathbb{N}_+$ referring to the number of affine-state constraints. Considering nonlinear systems for which such transformations exist might be regarded as rather restrictive, it is, however, a common assumption in nonlinear observer design. In fact, such a class of nonlinear systems is extensively studied in the literature.28-32 For instance, the transformation into system (2) includes the case of nonlinear state transformations in which a nonlinear dynamic system is affine in the unmeasured state $x_k$.

For the example of a single output and no input. The transformation of the nonlinear system allows obtaining system dynamics that are affine in the unmeasured state $x_k$ and, hence, an MHE formulation in which the underlying optimization problem is convex.17,20 More specifically, we design an MHE scheme in which the cost function consists of convex stage costs that penalize the output residuals and the model mismatch as well as a strongly convex Bregman distance function.17 For the sake of completeness and ease of analysis, we also center the Bregman distances associated with the model mismatch around zero. We handle the MHE problem by solving the following constrained optimization problem at each time instant.

**Estimation problem P:** At the time instant $k$, given the inputs $u_{k-N}, \ldots, u_{k-1}$, the measurements $y_{k-N}, \ldots, y_{k-1}$ and the a priori estimates $\tilde{z} = (\tilde{x}_{k-N}, 0)$, where $0 = (0, \ldots, 0)$ denotes $N$ zero vectors of dimension $n$ and

$$\tilde{x}_{k-N} = A (y_{k-N-1}) \hat{x}_{k-N-1} + \phi (y_{k-N-1}) + B (y_{k-N-1}) u_{k-N-1} + L (y_{k-N-1}) (y_{k-N-1} - C \hat{x}_{k-N-1}) , \hspace{1cm} (5a)$$

minimize the following cost

$$\sum_{i=k-N}^{k-1} q (\hat{w}_i) + r (\hat{y}_i - C \hat{x}_i) + D_\psi (\hat{z}, \tilde{z}) , \hspace{1cm} (5b)$$

subject to

$$\hat{x}_i = A (y_{i-1}) \hat{x}_{i-1} + \phi (y_{i-1}) + B (y_{i-1}) u_{i-1} + \hat{w}_{i-1}, \hspace{0.5cm} i = k - N + 1, \ldots, k - 1 \hspace{1cm} (5c)$$

$$C_x \hat{x}_i \leq d_x, \hspace{0.5cm} i = k - N, \ldots, k. \hspace{1cm} (5d)$$

We consider $\hat{z} = (\hat{x}_{k-N}, \hat{w})$ as the decision variable, where $\hat{w} = \{ \hat{w}_{k-N}, \ldots, \hat{w}_{k-1} \}$ refers to the model mismatch such that we can eliminate the predicted estimated states $\hat{x}_i$ in the optimization problem. Let $\hat{z}^* = (\hat{x}_{k-N}^*, \hat{w}^*)$ with $\hat{w}^* = \{ \hat{w}_{k-N}^*, \ldots, \hat{w}_{k-1}^* \}$ refer to the unique MHE estimate of (5), which yields the state estimate

$$\hat{x}_k = \prod_{i=k-N}^{k-1} A (y_i) \hat{x}_{k-N}^* + \sum_{i=k-N}^{k-1} \prod_{j=i+1}^{k-1} A (y_j) \left( \phi (y_j) + B (y_j) u_j + \hat{w}_j^* \right) , \hspace{1cm} (6)$$
and the estimation error

\[ e_{k-N} = x_{k-N} - \hat{x}_{k-N}^* . \]  

(7)

Here, \( D_\psi \) denotes a Bregman distance induced from a strongly convex function \( \psi \) and is defined as

\[ D_\psi (\hat{z}, \bar{z}) = \psi (\hat{z}) - \psi (\bar{z}) - (\hat{z} - \bar{z})^\top \nabla \psi (\bar{z}) . \]  

(8)

More detail on Bregman distances as well as some of their useful properties can be found in Appendix A1. The Bregman distance \( D_\psi \) acts in the proposed MHE scheme as a measure of the proximity between the estimated state \( \hat{x}_{k-N} \) and the stabilizing a priori estimate \( \bar{x}_{k-N} \) and is a generalization of the commonly used quadratic prior weighting. More importantly, its usage enables for the MHE scheme to inherit the stability properties offered by the a priori estimate. Moreover, it offers a rather flexible design that allows, for instance, eliminating the constraints (5d) and incorporate them into the cost function by choosing the so-called relaxed barrier functions as Bregman distances.20-22 From a technical point of view, and as we will see in the next section, the main advantage of the Bregman distance is that it enables to prove stability without an explicit representation of the error dynamics and independently of the rather general stage cost functions used in the optimization problem. Note that the proximity-based MHE formulation with convex stage cost and Bregman distances is inspired from the powerful class of proximal methods with Bregman distances. The basic proximal method iteratively finds a minimizer of a convex function by appending a Bregman distance centered at the current iterate and solving the resulting approximate problem instead. The proximal MHE formulation possesses a relatively similar structure. For this reason, we will adapt, in the following section, tools from the convergence analysis of proximal methods in order to be able to investigate the stability properties of the estimation error.

3 | STABILITY ANALYSIS

In this section, we investigate the stability properties of the estimation error of the MHE strategy (5). We impose the following assumptions.

Assumption 2. Suppose that the proposed MHE problem (5) satisfies the following assumptions.

A1: The stage costs \( r : \mathbb{R}^p \rightarrow \mathbb{R} \) and \( q : \mathbb{R}^n \rightarrow \mathbb{R} \) are convex, nonnegative, and achieve their corresponding minimum at zero.

A2: The Bregman distance \( D_\psi \) satisfies the following bounds for any \( \hat{z} = (\hat{x}, \hat{w}) \) and \( \bar{z} = (\bar{x}, \bar{w}) \)

\[ \frac{1}{2} \| \hat{x} - \bar{x} \|^2_{P_1} + c_1 \| \hat{w} - \bar{w} \|^2_2 \leq D_\psi (\hat{z}, \bar{z}) \leq \frac{1}{2} \| \hat{x} - \bar{x} \|^2_{P_2} + c_2 \| \hat{w} - \bar{w} \|^2_2 , \]  

(9)

for some \( 0 < P_1 \leq P_2 \in \mathbb{S}^n_{++} \) and \( 0 < c_1 \leq c_2 \).

Note that the strong convexity of \( D_\psi \) gives the quadratic lower bound in A2 of Assumption 2, while the upper bound comes from the additional assumption that the gradient of \( D_\psi \) is Lipschitz continuous. Even though the resulting class of Bregman distances might seem rather restrictive, it includes the important special case of quadratic distances \( D_\psi (x, y) = \| x - y \|^2_P \) induced from the function \( \psi(x) = \| x \|^2_P \), with \( P \) positive definite. In fact, quadratic distances are widely used in the MHE literature in order to replace the arrival cost, for which, typically, a closed form cannot be computed. In our setting, we can consider in addition any function of the form \( \psi(x) = \| x \|^2_P + B(x) \), where \( B : \mathbb{R}^n \rightarrow \mathbb{R} \) is a convex and strongly smooth function.

We will show that, if the weight matrices \( P_1 \) and \( P_2 \) from the quadratic bounds (9) satisfy a system of linear matrix inequalities (LMIs), then the estimation error (7) is guaranteed to be globally exponentially stable (GES). This will be proven by employing the Bregman distance \( D_\psi \) as a Lyapunov function. Before stating the first main result, we first need the following lemma that is obtained by analyzing problem (5) at some fixed time instant \( k \).
Lemma 1. Consider system (2) with the constraints (3) and the MHE scheme (5) at the time instant \( k \) and let \( A1 \) of Assumption 2 hold true. Then the following inequality holds

\[
D_\psi(z, \hat{z}) \leq D_\psi(z, \overline{z}),
\]

where \( \hat{z}^* = (\hat{x}_{k-N}^*, \hat{w}^*) \) refers to the unique solution of (5) and \( \overline{z} = (\overline{x}_{k-N}, \overline{0}) \) to the a priori estimates. Moreover, \( x_{k-N} \) in \( z = (x_{k-N}, \overline{0}) \) denotes the state of system (2) with zero model mismatch.

Proof. We rely in this proof on useful properties of the Bregman distances, which are summarized in Appendix A1. Let us first eliminate the forward prediction equations (5c) and rewrite the estimation problem in terms of the decision variables \( \hat{z} = (\hat{x}_{k-N}, \hat{w}) \). For given \( k \) and \( N \), let

\[
F_i(y) := \prod_{j=k-N}^{k-1} A_i(y_j) \in \mathbb{R}^{n \times n},
\]

\[
a_i(y) := \sum_{j=k-N}^{k-1} \prod_{l=j+1}^{i-1} A_j(y_l) (\phi(y_j) + B(y_j) u_j) \in \mathbb{R}^n.
\]

Hence, we can write down each state in the horizon window in terms of \( \hat{z} \) as follows

\[
\hat{x}_i = F_i(y) \hat{x}_{k-N} + a_i(y) + \sum_{j=k-N}^{i-1} \prod_{l=j+1}^{i-1} A_j(y_l) \hat{w}_j.
\]

We fix the time instant \( k \) and obtain for the estimation problem (5b), (5c), and (5d) the following condensed formulation

\[
\min_{\hat{z}} \quad f_k(\hat{z}) + D_\psi(\hat{z}, \overline{z}),
\]

s.t. \( \hat{z} \in S_k \),

where \( \hat{z} = (\hat{x}_{k-N}, \hat{w}) \),

\[
f_k(\hat{z}) := \sum_{i=k-N}^{k-1} q(\hat{w}_i) + r \left( y_i - CF_i(y) \hat{x}_{k-N} - C a_i(y) - C \sum_{j=k-N}^{i-1} \prod_{l=j+1}^{i-1} A_j(y_l) \hat{w}_j \right),
\]

accounts for the sum of stage costs and

\[
S_k := \left\{ (x, w) : C x F_i(y) x + C_x \sum_{j=k-N}^{i-1} \prod_{l=j+1}^{i-1} A_j(y_l) w_j \leq d_x - C_x a_i(y), \quad i = k - N, \ldots, k \right\},
\]

represents the constraint set associated with the reformulation of (5d) in terms of \( \hat{z} = (\hat{x}_{k-N}, \hat{w}) \). Since the cost of (13) is strongly convex, a unique solution \( \hat{z}^* \) exists. By optimality of \( \hat{z}^* \), it holds for all \( \hat{z} \in S_k \) that

\[
f_k(\hat{z}^*) + D_\psi(\hat{z}^*, \overline{z}) \leq f_k(\hat{z}) + D_\psi(\hat{z}, \overline{z}).
\]

Thus, for all \( \hat{z} \in S_k \) with \( f_k(\hat{z}) \leq f_k(\hat{z}^*) \), we deduce that

\[
D_\psi(\hat{z}^*, \overline{z}) \leq D_\psi(\hat{z}, \overline{z}).
\]

In other words, \( \hat{z}^* \) is the unique \( D_\psi \)-projection of \( \overline{z} \) onto the convex set \( \Omega_k \), that is,

\[
\hat{z}^* = \arg \min_{\hat{z} \in \Omega_k} D_\psi(\hat{z}, \overline{z}),
\]
where the set $\Omega_k$ is defined as

$$\Omega_k = \{ \hat{z} \in S_k : f_k(\hat{z}) \leq f_k(\check{z}) \}. \tag{19}$$

Since $z = (x_{k-N}, 0)$ yields

$$f_k(z) = \sum_{i=k-N}^{k-1} q(0) + r(y_i - CF_i y_{k-N} - C \alpha_i y') = \sum_{i=k-N}^{k-1} q(0) + r(0) = 0, \tag{20}$$

and lies in this polytopic set $S_k$, and since the stage costs $q$ and $r$ are nonnegative, it holds that $z \in \Omega_k$. Based on (A5) from Lemma 3 in the Appendix, we get

$$D_\psi(\check{z}^*, \bar{z}) \leq D_\psi(z, \bar{z}) - D_\psi(z, \check{z}^*). \tag{21}$$

Since the left-hand side of the above inequality is nonnegative, we obtain (10).

What we have shown is that, if we solve the optimization problem (5) at a given time instant $k$, the Bregman distance between the state of system (2) and the optimal solution is guaranteed to be smaller than the distance between this state and the stabilizing a priori estimate. On the basis of this Lemma, we are able to establish GES of the estimation error in the following theorem.

**Theorem 1.** Consider the proposed MHE scheme (5) for system (2) with the constraints (3) and let Assumption 2 hold true. If the weight matrices $P_1, P_2 \in S_{++}^{r+}$ from the quadratic bounds (9) on the Bregman distance satisfy the following LMI

$$(A(y_k) - L(y_k)C)^T P_2 (A(y_k) - L(y_k)C) - P_1 \preceq -Q, \tag{22}$$

for each output $y_k$ and some positive-definite matrix $Q$, then the dynamics of the estimation error (7) are GES.

**Proof.** We introduce in this proof the subindex $k$ to define the MHE errors at the time instant $k$. More precisely, the errors are given by $z_k - \hat{z}_k^* = (e_{k-N}, -\hat{w}_k^*)$, where $\hat{z}_k^* = (\check{x}_{k-N}^*, \hat{w}_k^*)$ and $\check{x}_{k-N}, 0)$ refer to the solution of the MHE problem (5) and the true state of system (2) with zero model mismatch, respectively. We prove this result by employing a Lyapunov function $V$, which satisfies the following conditions

$$\alpha_1 \|z_k - \check{z}_k^*\|^2 \leq V(z_k - \check{z}_k^*) \leq \alpha_2 \|z_k - \check{z}_k^*\|^2, \tag{23a}$$

and

$$\Delta V := V(z_k - \check{z}_k^*) - V(z_{k-1} - \check{z}_{k-1}^*) \leq -\alpha_3 \|z_{k-1} - \check{z}_{k-1}^*\|^2, \tag{23b}$$

for some positive constants $\alpha_1, \alpha_2$, and $\alpha_3$. We choose the Bregman distance in (5b) as a candidate Lyapunov function, that is,

$$V(z_k - \check{z}_k^*) = D_\psi(z_k, \check{z}_k^*). \tag{24}$$

Given the bounds (9) on $D_\psi$ in Assumption 2, the function $V$ satisfies (23a) with $\alpha_1 = \min(\lambda_{\text{min}}(P_1), c_1)/2$ and $\alpha_2 = \max(\lambda_{\text{max}}(P_2), c_2)/2$. Moreover, we have in view of (10) in Lemma 1 that

$$\Delta V = D_\psi(z_k, \check{z}_k^*) - D_\psi(z_{k-1}, \check{z}_{k-1}^*) \leq D_\psi(z_k, \check{z}_k) - D_\psi(z_{k-1}, \check{z}_{k-1}). \tag{25}$$

Using again the quadratic bounds (9), we obtain

$$\Delta V \leq \frac{1}{2} \|x_{k-N} - \check{x}_{k-N}\|^2_{P_2} + \frac{c_2}{2} \|Q - 0\|^2 - \frac{1}{2} \|x_{k-N-1} - \check{x}_{k-N-1}\|^2_{P_1} - \frac{c_1}{2} \|0 - \check{w}_{k-1}\|^2. \tag{26}$$
This is due to the fact that there is no model mismatch in (2) and that the associated a priori estimate equals zero. We show in the following that the stability of the MHE scheme can be inherited from the stabilizing a priori estimate \( \bar{x}_{k-N} \) by employing the LMI in (22). In fact, we get by using the dynamics of system (2) and the a priori estimate (5a)

\[
x_{k-N} - \bar{x}_{k-N} = (A(y_{k-N-1}) - L(y_{k-N-1})C)e_{k-N-1},
\]

since the input terms as well as the nonlinearities \( \phi \), which depend on the output, cancel out. Given that \( P_1 \) and \( P_2 \) satisfy (22), we obtain by substituting the latter equation in (26)

\[
\Delta V \leq \frac{1}{2} e_{k-N-1}^T (A(y_{k-N-1})^T P_2 A(y_{k-N-1}) - P_1) e_{k-N-1} - \frac{c_1}{2} \| \hat{w}_{k-1}^* \|^2
\]

\[
\leq -\frac{1}{2} \| e_{k-N-1} \|^2 + \frac{c_1}{2} \| \hat{w}_{k-1}^* \|^2,
\]

where \( A(y_{k-N-1}) = A(y_{k-N-1}) - L(y_{k-N-1})C \). Hence, the second condition (23b) is fulfilled with \( \alpha_3 = \min(\lambda_{\min}(Q), c_1)/2 \). This proves that the estimation error (7) is GES.

Observe that the LMI (22) implies that each resulting pair \( (A(y_k), C) \) is detectable for each \( y_k \in \mathbb{R}^p \). Moreover, the above result states that the existence of matrices \( P_1 \) and \( P_2 \), satisfying the LMI (22) for each output \( y_k \), is required to guarantee stability since arbitrary weight matrices might lead to instability as discussed later in Example 3. In general, finding these matrices in accordance to this condition and to Assumption A2 might be difficult. Nevertheless, we mention special cases for which this difficulty can be circumvented and simpler conditions can be achieved. First, one could check if the observer gain \( L(y_k) \) in (22) can be constructed such that \( A(y_k) - L(y_k)C \) is constant for all time instants \( k \) and does not depend on the output \( y_k \). This would reduce the problem to just one LMI that has to be fulfilled instead of a total of \( T - N + 1 \) LMIs, with \( T \) referring to the simulation time. Second, if \( A(y_k) \) is polynomial in \( y_k \) and the output trajectory is available offline, one could solve the LMI (22) for the weight matrices \( P_1 \) and \( P_2 \) using the sum of squares techniques. Third, we mention the case where the resulting system matrices in (2) do not depend on the output \( y_k \). This corresponds, for instance, to transformations of system (1) into the nonlinear observer form discussed earlier. We will study this case in the next section to show input-to-state stability of the proposed estimator.

**Remark 1.** It is worth pointing out that any stabilizing a priori estimate for which the Bregman distance constitutes a Lyapunov function can be used to ensure stability. For instance, one can construct a the a priori estimate based on the Kalman filter and choose the Bregman distance as a quadratic distance weighted with the inverse of the error covariance matrix.

### 4 | INPUT-TO-STATE STABILITY

In this section, we focus on a special case of transformations of system (1) to system (2) in which the system matrices \( A \) and \( B \) do not depend on the output \( y_k \). More specifically, consider

\[
x_{k+1} = Ax_k + \phi(\tilde{y}_k) + Bu_k + w_k,
\]

\[
y_k = Cx_k + v_k,
\]

where \( \tilde{y}_k : = C x_k \) refers to the nominal output and \( w_k \in \mathbb{R}^n \) and \( v_k \in \mathbb{R}^p \) denote the unknown process and measurement disturbances, respectively. We assume that the pair \( (A, C) \) is detectable and that \( x_k \) satisfies the constraints in (3). Moreover, we choose a quadratic prior weighting as a Bregman distance \( D_{\psi} \) in (5b), that is, \( D_\psi(x, \bar{x}) = \| x - \bar{x} \|^2_2 \), where \( P \in \mathbb{S}^{n+r} \). Note that \( w_k \) represents, in this case, an external disturbance that we do not consider in the optimization problem and that only the first state \( \hat{x}_{k-N} \) in the horizon window constitutes the decision variable. The MHE scheme for system (29) is given as follows.
**Estimation problem P**: At the time instant \( k \), given the inputs \( u_{k-N}, \ldots, u_{k-1} \), the measurements \( y_{k-N}, \ldots, y_{k-1} \), and the a priori estimate

\[
\bar{x}_{k-N} = A\hat{x}_{k-N-1} + \phi(y_{k-N-1}) + Bu_{k-N-1} + L(y_{k-N-1} - C\hat{x}_{k-N-1}) ,
\]  

minimizes the following cost

\[
\sum_{i=k-N}^{k-1} r(y_i - C\hat{x}_i) + \|\hat{x}_{k-N} - \bar{x}_{k-N}\|^2_P ,
\]  

subject to

\[
\hat{x}_i = A\hat{x}_{i-1} + \phi(y_{i-1}) + Bu_{i-1}, \quad i = k - N + 1, \ldots, k - 1
\]

\[
C_x \hat{x}_i \leq d_x, \quad i = k - N, \ldots, k
\]

with respect to \( \hat{x}_{k-N} \). Let \( \hat{x}_{k-N}^* \) refer to the unique MHE estimate of (30), which yields the state estimate

\[
\hat{x}_k = A^N \hat{x}_{k-N}^* + \sum_{i=k-N}^{k-1} A^{k-1-j} (\phi(y_i) + Bu_i)
\]

and the estimation error

\[
e_{k-N} = x_{k-N} - \hat{x}_{k-N}^* .
\]

In the following, we will show that, if the weight matrix \( P \) satisfies a simple LMI, then the estimation error is guaranteed to be input-to-state stable (ISS) with respect to the process and measurement disturbances.

**Assumption 3.** Suppose that the MHE problem (30) satisfies the following assumptions.

**A1**: The stage cost \( r : \mathbb{R}^p \to \mathbb{R} \) in (30b) is convex, nonnegative, and achieves its minimum at zero. Moreover, it is differentiable and its gradient is globally Lipschitz with parameter \( L_r \), that is,

\[
\|\nabla r(x) - \nabla r(y)\| \leq L_r \|x - y\| \quad \forall x, y \in \mathbb{R}^p .
\]

**A2**: The nonlinearity \( \phi \) is globally Lipschitz continuous with parameter \( L_\phi \), that is, we have

\[
\|\phi(x) - \phi(y)\| \leq L_\phi \|x - y\| \quad \forall x, y \in \mathbb{R}^p .
\]

**A3**: In the a priori estimate (30a), \( L \) is designed such that all the eigenvalues of \( A - LC \) are strictly within the unit circle.

Note that we need for the subsequent analysis the additional Lipschitz continuity assumption on the gradient of the stage cost \( r \) and on the nonlinearity \( \phi \), which are given in A1 and A2, respectively. As in the previous case, we present the following result.

**Lemma 2.** Consider the MHE scheme (30) for system (29) with the constraints (3) at the time instant \( k \) and let A1 of Assumption 3 hold true. Then,

\[
\|x_{k-N} - \hat{x}_{k-N}^*\|^2_P \leq \|x_{k-N} - \bar{x}_{k-N} - P^{-1} \sum_{i=k-N}^{k-1} O_i^T \nabla r \left( v_i + C \sum_{j=k-N}^{i-1} A^{i-j-1}w_j + \Delta \phi_j \right) \|^2_P ,
\]
where \( x_{k-N} \) denotes the state of (29), \( \bar{x}_{k-N} \) is the a priori estimate, and \( \Delta \phi_i := \phi(\bar{y}_i) - \phi(y_i) \).

**Proof.** Let us rewrite problem (30) in terms of the decision variable \( \hat{x}_{k-N} \). Let \( a_i = \sum_{j=k-N}^{i-1} A^{i-j-1} (\phi(y_j) + Bu_j) \) and \( \Theta_i := CA^{i-k+N} \) as well as

\[
G := \begin{bmatrix} C_x & C_x A & \vdots & C_x A^{N} \end{bmatrix} \in \mathbb{R}^{(N+1)q_k \times m}, \quad E := \begin{bmatrix} d_x - C_x a_{k-N+1}^{'} \\ \vdots \\ d_x - C_x a_k^{'} \end{bmatrix} \in \mathbb{R}^{(N+1)q_k}. 
\]

We reformulate the estimation problem (30b), (30c), and (30d) as follows

\[
\begin{align*}
\min_{\hat{x}_{k-N}} \quad & J_k (\hat{x}_{k-N}, \bar{x}_{k-N}) = \sum_{i=k-N}^{k-1} r (y_i - \Theta_i \hat{x}_{k-N} - C a_i) + \| \hat{x}_{k-N} - \bar{x}_{k-N} \|_p^2 \\
\text{s.t.} \quad & G \hat{x}_{k-N} \leq E.
\end{align*}
\]

The remainder of the proof constitutes of two parts. First, we show that, given the operator \( \rho : \mathbb{R}^n \rightarrow \mathbb{R}^n \) defined as:

\[
\rho(z) = \arg \min_{\hat{x}_{k-N}} J_k (\hat{x}_{k-N}, z) \tag{37a}
\]

\[
\text{s.t.} \quad G \hat{x}_{k-N} \leq E, \tag{37b}
\]

we can derive for a given time instant \( k \) a nonexpansiveness property of \( \rho \) as developed for proximal point algorithms.\(^{25,35}\)

Second, we use this property to obtain the result of the lemma.

**Part 1.** Since the overall cost function is strongly convex, the minimizer \( \rho(z) \) is unique for each \( z \). Hence, if \( z' \neq z \), it holds that \( \rho(z') \neq \rho(z) \). We derive the Karush-Kuhn-Tucker (KKT) optimality conditions corresponding to problem (37) at the minimum \( \rho(z) \) for an a priori estimate \( z \). We obtain

\[
0 = P (\rho(z) - z) - \sum_{i=k-N}^{k-1} \Theta_i^\top \nabla r (y_i - \Theta_i \rho(z) - C a_i) + G^\top \lambda
\]

\[
\lambda^\top (G \rho(z) - E) = 0
\]

\[
G \rho(z) - E \leq 0, \quad \lambda \geq 0,
\]

where \( \lambda \in \mathbb{R}^{(N+1)q_k} \) denotes the associated Lagrange multiplier. For an a priori estimate \( z' \), we can write down the corresponding KKT conditions at the minimum \( \rho(z') \) with the Lagrange multiplier \( \lambda' \). Subtracting the gradient condition (38a) at \( \rho(z') \) from the condition at \( \rho(z) \) and then multiplying from the left by \( (\rho(z) - \rho(z'))^\top \) yields

\[
(\rho(z) - \rho(z'))^\top P (z - z') = \| \rho(z) - \rho(z') \|_p^2 + (\rho(z) - \rho(z'))^\top G^\top (\lambda - \lambda')
\]

\[
+ \sum_{i=k-N}^{k-1} - (\rho(z) - \rho(z'))^\top \Theta_i^\top \left[ \nabla r (y_i - \Theta_i \rho(z) - C a_i) - \nabla r (y_i - \Theta_i \rho(z') - C a_i) \right].
\]

By defining \( \bar{y}_i = y_i - \Theta_i \rho(z) - C a_i \) and \( \bar{y}_i' = y_i - \Theta_i \rho(z') - C a_i \), each \( i \)th term of the sum in the latter equality can be computed as follows

\[
(\bar{y}_i - \bar{y}_i')^\top (\nabla r (\bar{y}_i) - \nabla r (\bar{y}_i')), \quad i = k - N, \ldots, k - 1.
\]

Since the operator \( \nabla r \) is convex, we know that (40) is nonnegative. Hence, we obtain in view of (39) the following inequality

\[
(\rho(z) - \rho(z'))^\top P (z - z') \geq \| \rho(z) - \rho(z') \|_p^2 + [G(\rho(z) - \rho(z'))]^\top (\lambda - \lambda').
\]
Given the complementary slackness condition (38b), the second term of the right-hand side of (41) can be written as

$$
\left( \lambda^T G \rho(z) - \lambda^T G \rho(z') \right)^T + \left( \lambda^T G \rho(z') - \lambda^T G \rho(z) \right)^T = \left( \lambda^T E - \lambda^T G \rho(z') \right)^T + \left( \lambda^T E - \lambda^T G \rho(z) \right)^T
$$

$$= (E - G \rho(z'))^T \lambda + (E - G \rho(z))^T \lambda' \geq 0,
$$

(42)

where the fact that the latter expression is nonnegative holds as a result of (38c). Hence, (41) becomes

$$\| \rho(z) - \rho(z') \|^2_P \leq (\rho(z) - \rho(z'))^T P (z - z')$$

$$= (\rho(z) - \rho(z'))^T P^{1/2} P^{1/2} (z - z').
$$

(43)

Since $P^{1/2}$ is also symmetric positive definite, we obtain in view of the Cauchy-Schwarz inequality

$$\| \rho(z) - \rho(z') \|^2_P \leq \|P^{1/2} (\rho(z) - \rho(z'))\| \|P^{1/2} (z - z')\|.
$$

(44)

By taking the square of (44), we get

$$\| \rho(z) - \rho(z') \|^4_P \leq \|P^{1/2} (\rho(z) - \rho(z'))\|^2 \|P^{1/2} (z - z')\|^2
$$

$$= \| \rho(z) - \rho(z') \|^2_P \| z - z' \|^2_P.
$$

(45)

If $z \neq z'$, then $\rho(z) \neq \rho(z')$ and we can divide by $\| \rho(z) - \rho(z') \|^2_P$ to get the following nonexpansiveness property

$$\| \rho(z) - \rho(z') \|^2_P \leq \| z - z' \|^2_P,
$$

(46)

for the operator $\rho$ defined in (37). Otherwise, $\rho(z) = \rho(z')$ and we would get $0 \leq 0$ in (45).

**Part 2.** The idea is to choose $z$ such that $p(z) = x_{k-N}$ in order to employ the nonexpansiveness property (46) for the subsequent Lyapunov analysis. Substituting the true state $x_{k-N} = p(z)$ of system (29) into (38a) yields

$$0 = P(x_{k-N} - z) - \sum_{i=k-N}^{k-1} \mathcal{O}_i^T \nabla r(y_i - \mathcal{O}_i x_{k-N} - \mathcal{C} a_i).
$$

(47)

This holds since the true state $x_{k-N}$ strictly satisfies the constraints (36a), which yields in view of the complementary slackness condition (38b) that the associated Lagrange multiplier equals zero. The true measurements in (29) are given by

$$y_i = \mathcal{O}_i x_{k-N} + C \tilde{a}_i + v_i + C \sum_{j=k-N}^{i-1} A^{i-j-1} w_j,
$$

(48)

where $\tilde{a}_i = \sum_{j=k-N}^{i-1} A^{i-j-1} (\phi (\tilde{y}_j) + B u_j)$. Substituting the latter expression in (47) yields

$$z = x_{k-N} - P^{-1} \sum_{i=k-N}^{k-1} \mathcal{O}_i^T \nabla r \left( v_i + C \sum_{j=k-N}^{i-1} A^{i-j-1} w_j + \left( \phi(\tilde{y}_i) - \phi (y_i) \right) \right),
$$

(49)

since $P$ is invertible. Hence, evaluating the property (46) at $z$ computed in (49) with $p(z) = x_{k-N}$, and $z' = \bar{x}_{k-N}$ with $p(z') = \bar{x}_{k-N}^+$ yields the desired inequality (35).

If we set the disturbances to zero, the result (35) can be regarded as a special case of (10) with $D_p(x, \bar{x}) = \|x - \bar{x}\|^2_P$. However, observe that the corresponding proofs are quite different. This is due to the fact that, in contrast to the proof of Lemma 1, the presence of disturbances in system (29) did not allow us to take advantage of the central property of Bregman distances introduced in Lemma 3 in Appendix A1. More precisely, we needed in Lemma 1 the crucial step that the cost function evaluated at the true state achieves its minimal value of zero, which does not hold in the perturbed case. We establish in the following theorem that the estimation error is ISS with respect to the disturbances.
Theorem 2. Consider the proposed MHE scheme (30) for system (29) with constraints (3) such that Assumption 3 holds true. If we choose the weight matrix \( P \in S_{n+}^+ \) in (30b) such that it satisfies the following LMI

\[
(A - LC)^T P (A - LC) - P \leq -Q,
\]

for some positive-definite matrix \( Q \in S_{n+}^+ \), then the corresponding estimation error dynamics are ISS with respect to the process and measurement disturbances.

Proof. Let \( p := |P|, \ \overline{p} := |A|, \ \overline{a} := |A - LC|, \ \overline{b} := |C|, \ \overline{c} := |L|, \ \overline{\rho} := |\Theta|, \ \overline{\sigma} := |\gamma|, \ p = \lambda_{\text{min}}(P), \) and \( q = \lambda_{\text{min}}(Q) \). Following the lines of the proof of Theorem 1, we choose the employed Bregman distance as a candidate Lyapunov function, that is, \( V(e_{k-N}) = \| e_{k-N} \|_P \). The conditions on \( V \) such that it can constitute an ISS-Lyapunov function are given in Definition 1 in Appendix B1. Note that condition (B3a) is already fulfilled. Moreover, we have in view of (35)

\[
V(e_{k-N}) - V(e_{k-N-1}) = \| e_{k-N} \|_P^2 - \| e_{k-N-1} \|_P^2
\]

Substituting the state of system (29) and the a priori estimate (30a) yields

\[
x_{k-N} - \overline{x}_{k-N} = Ax_{k-N-1} + \phi(y_{k-N-1}) + w_{k-N-1} - Ax_{k-N-1} - \overline{\phi}(y_{k-N-1}) - L(y_{k-N-1} - Cx_{k-N-1})
\]

\[
= Ax_{k-N-1} + w_{k-N-1} + \overline{\phi}k_{k-N-1} - Ax_{k-N-1} - \overline{\phi}k_{k-N-1} - L(Cx_{k-N-1} + v_{k-N-1} - Cx_{k-N-1})
\]

\[
= (A - LC)e_{k-N-1} + \overline{\phi}k_{k-N-1} + w_{k-N-1} - Lv_{k-N-1}.
\]

Thus, we get

\[
V(e_{k-N}) - V(e_{k-N-1}) \leq \| (A - LC)e_{k-N-1} + d_{k-N-1} \|_P^2 - \| e_{k-N-1} \|_P^2
\]

\[
= e_{k-N-1}^T ((A - LC)^T P (A - LC) - P) e_{k-N-1} + 2 d_{k-N-1}^T P(A - LC)e_{k-N-1} + d_{k-N-1}^T P d_{k-N-1},
\]

where

\[
d_{k-N-1} := w_{k-N-1} - Lv_{k-N-1} - P^{-1} \sum_{i=k-N}^{k-1} \Theta^T_i \gamma_i (v_i + C \sum_{j=k-N}^{i-1} A^{i-j-1} w_j + \Delta \phi_j) + \overline{\phi}k_{k-N-1},
\]

accounts for the noise terms. Exploiting the LMI in (50), we get for the first term in the right-hand side of (53)

\[
e_{k-N-1}^T ((A - LC)^T P (A - LC) - P) e_{k-N-1} \leq -q\| e_{k-N-1} \|^2.
\]

By using Young’s inequality, we obtain for the second term

\[
2 d_{k-N-1}^T P(A - LC) e_{k-N-1} \leq s\| e_{k-N-1} \|^2 + \frac{\overline{a} \overline{p}}{s} \| d_{k-N-1} \|^2,
\]

where \( s \in \mathbb{R}_{++} \) is arbitrary. For the third term, we have

\[
d_{k-N-1}^T P d_{k-N-1} \leq \overline{p} \| d_{k-N-1} \|^2.
\]

Hence, we finally obtain

\[
V(e_{k-N}) - V(e_{k-N-1}) \leq -q(s - s)\| e_{k-N-1} \|^2 + \overline{p}(1 + \overline{a}/s) \| d_{k-N-1} \|^2
\]

\[
= -c_1\| e_{k-N-1} \|^2 + c_2\| d_{k-N-1} \|^2,
\]
where \( c_1 := q - s \) and \( c_2 := \overline{p}(1 + \overline{a}/s) \) are strictly positive if we choose \( s \) such that \( 0 < s < q \). Thus, condition (B3b) is fulfilled and \( \overline{V} \) is an ISS-Lyapunov function for the estimation error with respect to the disturbances in \( d_{k-N-1} \) defined in (54). For the sake of completeness, we also compute the ISS gains. The computations can be found in Appendix C1. In particular, we show that there exist \( \beta_e \in (0, 1) \) and \( c_e, c_w, c_v \in \mathbb{R}_{++} \) such that for all \( \xi = e_0 \in \mathbb{R}^n \) and any \( k \)

\[ \|e_{k-N}\| \leq c_e \beta^k e^{-N} \|\xi\| + c_w \|w_{[k-1]}\| + c_v \|v_{[k-1]}\|, \]

with

\[ c_w = 1 + \frac{L_r}{P} \bar{c}^2 \sum_{i=1}^{N-1} \bar{a} \sum_{j=0}^{i-1} \bar{a}, \quad c_v = \bar{i} + L_\phi + \frac{L_r}{P} \bar{c} \sum_{i=0}^{N-1} \bar{a} (1 + \bar{i}L_\phi). \]

\[ 5 \mid \text{SIMULATION EXAMPLES} \]

We illustrate the stability results obtained in the previous two sections by means of three simulations examples. More precisely, we demonstrate the global exponential stability established in Theorem 1 in the first example and input-to-state stability established in Theorem 2 in the second and third examples.

\[ 5.1 \mid \text{Example 1} \]

In this section, we consider an example of nonlinear systems in which states are multiplied with a (potentially nonlinear) function of states that can be measured. By reformulating these nonlinearities in terms of the outputs, we can obtain a system of form (2). For the sake of illustration, we consider a simple nonlinear system of the form (1) where

\begin{equation}
\begin{bmatrix}
\zeta^{(1)}_{k+1} \\
\zeta^{(2)}_{k+1}
\end{bmatrix} = \begin{bmatrix}
z^{(1)}_{k} \\
z^{(2)}_{k}
\end{bmatrix} = f(z_k),
\end{equation}

\[ y_k = z^{(1)}_k = h(z_k), \]

which can be transformed to system (2) as follows

\begin{equation}
\begin{aligned}
x_{k+1} &= \begin{bmatrix} 0 & 1 \\ 0 & y_k \end{bmatrix} x_k = A(y_k) x_k, \\
y_k &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_k = C x_k.
\end{aligned}
\end{equation}

Observe that the state \( x_k \) obtained from this transformation is equivalent to the state \( z_k \) of system (61), that is, the evolutions of both systems are identical. Also note that the new system can be viewed as a discrete-time linear time varying system where \( A(y_k) \) varies at each time instant \( k \) based on an a given output trajectory. We estimate the state \( x_k \) of system (62) by solving the MHE problem (5) with horizon length \( N = 2 \) and stage costs \( q(x) = 0 \) and \( r(x) = \|x\| \). Moreover, we choose the Bregman distance \( D_\psi \) in the cost function (5b) as \( D_\psi = \|x - \tilde{x}\|_p^2 \) where the weight matrix

\[ P = \begin{bmatrix}
0.0470 & -0.0380 \\
-0.0380 & 0.0491
\end{bmatrix}, \]

\[ \text{Remark 2.} \] Note that an ISS result with matrices \( A \) and \( B \) depending on the output might be difficult to achieve. This is due to the fact that, in the estimation problem, the measurement noise will also enter the system matrix \( A \). More specifically, in the Lyapunov analysis, we will not be able to separate the disturbances from the estimation error in order to arrive to the desired ISS property of the form (58).
FIGURE 1 Evolution of the state of system (61) and of the state estimates resulting from employing the proposed moving horizon estimation (MHE) scheme (5) as well as the Luenberger observer (64) [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 2 The resulting estimation errors for the Luenberger observer and the proposed estimation strategy (5) [Colour figure can be viewed at wileyonlinelibrary.com]

satisfies the LMIs (22) with \( P = P_1 = P_2 \). We compare the results of the MHE scheme (5) with those obtained by employing a Luenberger observer. For a fair comparison, we design the Luenberger observer with the same \( L \) matrices as in (5a), which gives the following observer dynamics

\[
\dot{\hat{x}}_k = A (y_{k-1}) x_{k-1} + L (y_{k-1}) (y_{k-1} - C \hat{x}_{k-1}).
\] (64)

We set the initial condition \( z_0 = x_0 = [0.51.5]^T \) and choose the initial guess \( \hat{x}_0 = [01.9]^T \) for both estimators. The state estimation results are presented in Figure 1. The resulting estimation errors for both estimators are shown in Figure 2. We can see that the proximity MHE scheme (5) reconstructs the states of the nonlinear system (61) and exhibits a good estimation performance. In fact, its performance is superior to the one of the Luenberger observer even with a small horizon length of 2. This observation was rather expected since the proposed MHE strategy employs the Luenberger observer as an a priori estimate and performs an additional optimization based on the available measurements, that is, the MHE scheme should perform at least as good as the Luenberger observer designed accordingly.

5.2 Example 2

We consider the following discrete-time nonlinear system:\(^{28}\)

\[
\begin{align*}
 z_{k+1} &= \begin{bmatrix} z_{k+1}^{(1)} \\ z_{k+1}^{(2)} \\ z_{k+1}^{(3)} \\ z_{k+1}^{(4)} \\ z_{k+1} \end{bmatrix} = \begin{bmatrix} z_k^{(1)} \\ z_k^{(2)} \\ z_k^{(3)} \\ z_k^{(4)} + \left( z_k^{(3)} \right)^2 \\ z_k + \left( z_k^{(2)} \right)^2 \end{bmatrix} = f(z_k), \\
 y_k &= \begin{bmatrix} y_k^{(1)} \\ y_k^{(2)} \end{bmatrix} = \begin{bmatrix} z_k^{(3)} \\ z_k^{(3)} + \left( z_k^{(2)} \right)^2 \end{bmatrix} = h(z_k),
\end{align*}
\] (65a)
FIGURE 3  The evolutions of the second and third states of system (29), (66) along with the resulting state estimates when using the Luenberger observer and the proximal moving horizon estimation (MHE) scheme (30) [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 4  The resulting estimation errors for the Luenberger observer and the proximal moving horizon estimation (MHE) scheme (30) designed for the perturbed system (29), (66) [Colour figure can be viewed at wileyonlinelibrary.com]

with initial condition $z_0 = [1, 0.5, 0.3, 0.15]^{\top}$. This system can be transformed via the state transformation $x = T(z) = \begin{bmatrix} z^{(1)} , z^{(2)} , z^{(3)} , (z^{(2)})^2 \end{bmatrix}^{\top}$ into a system in the nonlinear observer form given by (29) with

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} , \quad C = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} , \quad \phi(y_k) = \begin{bmatrix} y^{(1)}_k \ 0 \\ 0 \ y^{(2)}_k \end{bmatrix} , \quad B = 0 , \quad (66)$$

zero disturbances and initial condition $x_0 = T(z_0) = [1, 0.5, 0.3, 0.4]^{\top}$. Note that the resulting pair $(A, C)$ is observable. We design for the new system the proposed proximity MHE scheme (30) with horizon length $N = 3$ and stage cost $r(x) = R \|x\|^2$ with $R = 0.01$. We choose the weight matrix $P$ in (30b) such that the LMI (50) is satisfied for the designed observer gain $L$. In order to illustrate the established ISS results, we assume that the new system is subject to process and measurement disturbances, which are chosen as independent, zero mean, Gaussian random variables with variances $\sigma_w^2 = 0.01, 1, 1, 1$ and $\sigma_v^2 = 0.1, 1, 1$, respectively. As was done in the previous case, we compare the results of the proximal MHE scheme (30) with a Luenberger observer design with the same observer gain $L$ and initial guess $\hat{x}_0 = [0.3, 0.1, 0.4, 0.2]^{\top}$. The estimation results are depicted in Figures 3 and 4. We can see that, in the presence of process and measurement disturbances, the proximal MHE approach yields ISS estimation errors. Moreover, the estimated states converge to a neighborhood of the true state of system (29), (66). This is also the case for the Luenberger observer. However, observe that the proximal MHE scheme with a quadratic stage cost exhibits a better estimation performance and is more robust in the presence of the Gaussian noise.

5.3  Example 3

In this section, we focus on an example of nonlinear systems in which every nonlinearity can be expressed in terms of the output. This yields a system with a constant matrix $A$, where all nonlinear terms are stacked into the nonlinearity $\phi(y_k)$. [diagram of nonlinearity function]
More specifically, we consider the example of a flexible joint robot link where joint flexibility is modeled by a stiffening torsional spring. The associated continuous-time nonlinear model is given by

\[
\begin{aligned}
\dot{\theta}_m &= \omega_m \\
\dot{\omega}_m &= \frac{1}{J_m} \tau - \frac{b}{J_m} \omega_m + \frac{K_r}{J_m} u \\
\dot{\theta}_l &= \omega_l \\
\dot{\omega}_l &= -\frac{1}{J_l} \tau - \frac{Mgh}{J_l} \sin(\theta_l),
\end{aligned}
\]  

(67)

where \(\theta_m, \omega_m, \theta_l, \) and \(\omega_l\) denote the motor and link positions and velocities, respectively, and \(\tau\) is the torque, which is computed as

\[
\tau = \kappa_1 (\theta_l - \theta_m) + \kappa_2 (\theta_l - \theta_m)^3,
\]

(68)

with \(\kappa_1 = \kappa_2 = 1\) due to the stiffening spring. The parameters of the model are presented in Table 1. The input is given by \(u(t) = \sin(t)\) and we assume to measure the motor and link positions, that is, \(y = [y_1 \ y_2]^T = [\theta_m \ \theta_l]^T\). We define the state vector \(x = [\theta_m \ \omega_m \ \theta_l \ \omega_l]^T\) and transform system (67) into the following form

\[
\begin{aligned}
\dot{x} &= A_c \ x + \phi_c(y) + B_c \ u, \\
y &= C \ x,
\end{aligned}
\]  

(69a)

(69b)

where

\[
A_c = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\frac{b}{J_m} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \frac{K_r}{J_m} \\ 0 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \phi_c(y) = \begin{bmatrix} 0 \\ \frac{\kappa_1}{J_m} (y_2 - y_1) + \frac{\kappa_2}{J_m} (y_2 - y_1)^3 \\ 0 \\ -\frac{\kappa_1}{J_l} (y_2 - y_1) - \frac{\kappa_2}{J_l} (y_2 - y_1)^3 - \frac{Mgh}{J_l} \sin(y_2) \end{bmatrix}.
\]

(69c)

We discretize system (69) using the Euler method with a sampling time of \(h = 0.01\) and obtain a discrete-time system of the form (29), where \(A = I_4 + hA_c, \ B = hB_c, \) and \(\phi(y_k) = h \phi_c(y_k)\). Here, the pair \((A, C)\) is observable and the initial condition is set to \(x_0 = [0.250 \ 0.30 \ 0.15]^T\). We design for the resulting system two proximal MHE schemes (30), both with horizon length of \(N = 5\), initial guess \(\hat{x}_0 = [0.1 \ 0.1 \ 0.1]^T\), observer gain \(L\) such that the eigenvalues of \(A - LC\) are \([0.90 \ 0.40 \ 0.5]^T\), and weight matrix \(P\) such that the LMI (50) is satisfied but with different stage costs \(r_{e_i}(x) = \|x\|^2\) and \(r_{e_i}(x) = \|x\|_1\). We investigate the performance and robustness of the estimators by assuming that the measurements are corrupted by some additive noise, that is, \(y = Cx + v\). In particular, we consider the two following scenarios: (a) \(v\) consists of outliers, more specifically \(v\) is zero everywhere except for \(v(t_1) = 5\) and \(v(t_2) = 7\), where \(t_1 = 0.2\) second and \(t_2 = 0.5\) second and (b) \(v = z + e\) where \(z\) is a zero mean Gaussian noise with standard deviation \(\sigma_e = 0.01[1, 1]^T\) and \(e = 3\) is an.

**TABLE 1** Parameters of the model of a flexible joint robot link

| Symbol | Description | Value | Units |
|--------|-------------|-------|-------|
| \(J_m\) | Inertia of the motor | \(3.7 \times 10^{-3}\) | kg m\(^2\) |
| \(J_l\) | Inertia of the link | \(9.3 \times 10^{-3}\) | kg m\(^2\) |
| \(2h\) | Length of the link | 0.15 | m |
| \(M\) | Mass of the link | 0.21 | kg |
| \(b\) | Viscous friction | 0.046 | m |
| \(K_r\) | Amplifier gain | 0.08 | N m V\(^{-1}\) |
The resulting estimation error dynamics for two proximal moving horizon estimation (MHE) schemes (30) with different stage costs designed for the discretized system of (69) and scenario (a), where the measurements are corrupted by two outliers. The estimation errors of the Luenberger observer with the same observer gain $L$ are also depicted [Colour figure can be viewed at wileyonlinelibrary.com]

The resulting estimation errors for two proximal moving horizon estimation (MHE) schemes (30) with different stage costs designed for the discretized system of (69) and scenario (b), where the measurements are corrupted by Gaussian noise and an outlier [Colour figure can be viewed at wileyonlinelibrary.com]

| $N$  | 1     | 2     | 5     | 10    | 20    | 30    |
|------|-------|-------|-------|-------|-------|-------|
| pMHE ($\ell_2$) | 1.5718 | 1.0259 | 0.5720 | 0.7592 | 0.4362 | 0.2327 |
| pMHE ($\ell_1$) | 1.6647 | 0.3031 | 0.1300 | 0.0048 | $2.8916 \times 10^{-4}$ | $1.1006 \times 10^{-4}$ |

Abbreviations: MHE, moving horizon estimation; pMHE, proximal MHE; RMSE, root-mean-square error.

Figure 5 shows that both proximal MHE schemes (denoted by pMHE) exhibit a much better performance than the Luenberger observer. Nevertheless, we can see that the MHE scheme with a quadratic stage cost is affected by the outliers and shows rather large estimation errors. In contrast to this behavior, observe that the estimator with the $\ell_1$-norm as a stage cost is insensitive to the outliers and that the corresponding estimation errors are GES. The results of scenario (b) in Figure 6 confirm this observation and highlight the performance of both estimators with different noise characteristics. More precisely, in the presence of Gaussian noise, the MHE scheme with the $\ell_2$-norm performs in this example much better than the MHE scheme with the $\ell_1$-norm. However, when the outlier takes place, the estimation errors of the quadratic estimator become larger than those of the pMHE with the $\ell_1$-norm. These observations are validated by the computed root-mean-square errors (RMSEs)

$$\text{RMSE} = \sqrt{\frac{1}{T_{\text{sim}} - N + 1} \sum_{k=N}^{T_{\text{sim}}} \|e_k\|^2},$$

associated with each proximal MHE scheme, where $T_{\text{sim}} = 100$ denotes the simulation time for system (69). These are reported in Tables 2 and 3 for scenarios (a) and (b), respectively, along with the performance of the MHE schemes with other values of the horizon length $N$.

In scenario (a), since the MHE scheme with the $\ell_1$-norm is robust against outliers and exhibits GES of the estimation errors, its performance improves with an increasing value of the horizon length $N$, as depicted in Table 2. In fact, for $N = 30$, its RMSE is much smaller than the one corresponding to the MHE scheme with the $\ell_2$-norm. In scenario (b), we can see in Table 3 that both schemes achieve a better performance with a larger horizon length. Moreover, their
TABLE 3  RMSE of the employed proximal MHE schemes for scenario (b) and different values of N

| N   | 1    | 2    | 5    | 10   | 20   | 30   |
|-----|------|------|------|------|------|------|
| pMHE ($\ell_2$) | 0.6212 | 0.4572 | 0.3348 | 0.3260 | 0.1808 | 0.1086 |
| pMHE ($\ell_1$) | 0.6665 | 0.3330 | 0.2556 | 0.1089 | 0.0341 | 0.0222 |

Abbreviations: MHE, moving horizon estimation; pMHE, proximal MHE; RMSE, root-mean-square error.

TABLE 4  Mean CPU time (in millisecond) per iteration of the employed estimators

| Scenario (a) | Scenario (b) |
|--------------|--------------|
| Luenberger observer | 0.0434 | 0.0448 |
| pMHE ($\ell_2$) | 21.7197 | 21.1148 |
| pMHE ($\ell_1$) | 30.7448 | 31.9686 |

Abbreviation: pMHE, proximal moving horizon estimation.

FIGURE 7  The resulting estimation errors for the proximal MHE schemes (30) compared to those of the EKF, where the measurements are corrupted by outliers as described in scenario (a) [Colour figure can be viewed at wileyonlinelibrary.com]

performances are more comparable, which could be explained by the fact that, even though the $\ell_1$-norm is more robust against the presence of the outlier, the $\ell_2$-norm is more suitable when white noise is present.

Table 4 reports the required computational effort of the proximal MHE schemes in comparison with the Luenberger observer. Obviously, the least computationally involved scheme is the Luenberger observer. Since the pMHE problem with the $\ell_2$-norm constitutes a quadratic program, we can observe that its solution requires less time than the pMHE scheme with the $\ell_1$-norm. Note that both optimization problems are solved with the same nonlinear programming solver of Matlab.

For scenario (a), we also compare the estimation results with those provided from the extended Kalman filter (EKF). The EKF estimates the state of the nonlinear system obtained after discretizing system (67) using the Euler method with a sampling time of $h = 0.01$. Note that the tuning of the process and measurement noise matrices $Q_{ekf}$ and $R_{ekf}$ is by no means trivial. More specifically, $R_{ekf}$ has to be chosen large enough such that the influence of the outliers can be attenuated, but also not so large such that the information provided from the measurements becomes neglected. After tuning, we obtain the results depicted in Figure 7, in which we can observe that the EKF is highly affected by the presence of outliers and that both MHE schemes exhibit a much better performance.

It is important to mention that, if we set the weight matrix $P$ as a free-tuning parameter and choose it as the identity matrix for instance, the estimation error would diverge in this example. In fact, this specific choice of $P$ does not satisfy the LMI (50). This demonstrates that, even though the proximal MHE scheme is based on a stabilizing Luenberger observer, the weights have to be chosen accordingly, for example, such that the LMI (22) is fulfilled for the generalized Bregman case or such that the LMI (50) is satisfied for the quadratic case in order to guarantee stability.

6  |  CONCLUSION

In this article, we presented a proximity-based MHE formulation for a class of discrete-time nonlinear systems, which can be transformed into systems that are affine in the unmeasured state. The convex cost function of the estimation problem consists of two terms. The first term specifies the performance of the estimator and comprises the sum of convex stage costs, which take the recent online measurements into account. The second term is designed to ensure stability and
consists of Bregman distances that penalize the deviation from a stabilizing a priori estimate. As a first main result, we showed the global exponential stability of the resulting estimation error. For a special case that includes the class of linear systems with additive disturbances, we established as a second main result the input-to-state stability of the underlying estimation error with respect to process and measurement disturbances under rather mild assumptions.

Future research can focus on deriving weaker conditions for stability and to modify the design of the estimator accordingly. While the imposed system class seems to be rather restrictive, it enables as a future work to investigate employing an EKF as an a priori estimate instead of the Luenberger observer by replacing the output dependent matrices in (2) by time-varying ones. This might have the potential benefit of enhancing the local stability and performance properties of the widely used EKF.

**ORCID**

Meriem Gharbi https://orcid.org/0000-0002-0466-5101

**REFERENCES**

1. Rawlings JB, Mayne DQ, Diehl M. Model Predictive Control: Theory, Computation, and Design. Madison, WI: Nob Hill Publishing; 2017.
2. Alessandri A, Baglietto M, Battistelli G. Receding-horizon estimation for discrete-time linear systems. *IEEE Trans Autom Control*. 2003;48(3):473-478.
3. Rao CV, Rawlings JB, Lee JH. Constrained linear state estimation—a moving horizon approach. *Automatica*. 2001;37(10):1619-1628.
4. Sui D, Johansen TA. Linear constrained moving horizon estimator with pre-estimating observer. *Syst Control Lett.* 2014;67:40-45.
5. Rao CV, Rawlings JB, Mayne DQ. Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Trans Autom Control*. 2003;48(2):246-258.
6. Alessandri A, Baglietto M, Battistelli G. Moving-horizon state estimation for nonlinear discrete-time systems: New stability results and approximation schemes. *Automatica*. 2008;44(7):1753-1765.
7. Haber A, Verhaegen M. Moving horizon estimation for large-scale interconnected systems. *IEEE Trans Autom Control*. 2013;58(11):2834-2847.
8. Schneider R, Hanne mann-Tamás R, Marquardt W. An iterative partition-based moving horizon estimator with coupled inequality constraints. *Automatica*. 2015;61:302-307.
9. Zhang J, Liu J. Distributed moving horizon state estimation for nonlinear systems with bounded uncertainties. *J Process Control*. 2013;23(9):1281-1295.
10. Alessandri A, Baglietto M, Battistelli G. Min-max moving-horizon estimation for uncertain discrete-time linear systems. *SIAM J Control Optim*. 2012;50(3):1439-1465.
11. Alessandri A, Awawdeh M. Moving-horizon estimation with guaranteed robustness for discrete-time linear systems and measurements subject to outliers. *Automatica*. 2016;67:85-93.
12. Liu J. Moving horizon state estimation for nonlinear systems with bounded uncertainties. *Chem Eng Sci*. 2013;93:376-386.
13. Müller MA. Nonlinear moving horizon estimation in the presence of bounded disturbances. *Automatica*. 2017;79:306-314.
14. Wynn A, Vukov M, Diehl M. Convergence guarantees for moving horizon estimation based on the real-time iteration scheme. *IEEE Trans Autom Control*. 2014;59(8):2215-2221.
15. Alessandri A, Gaggero M. Fast moving horizon state estimation for discrete-time systems using single and multi iteration descent methods. *IEEE Trans Autom Control*. 2017;62(9):4499-4511.
16. Kong H, Sukkarieh S. Suboptimal receding horizon estimation via noise blocking. *Automatica*. 2018;98:66-75.
17. Gharbi M, Ebenbauer C. A proximity approach to linear moving Horizon estimation. *IFAC-PapersOnLine*. 2018;51(20):549-555.
18. Donoho DL. Compressed sensing. *IEEE Trans Inf Theory*. 2006;52(4):1289-1306.
19. Candes EJ, Wakin MB, Boyd SP. Enhancing sparsity by reweighted $\ell_1$ minimization. *J Fourier Anal Appl*. 2008;14(5-6):877-905.
20. Gharbi Meriem, Ebenbauer Christian. A proximity moving Horizon estimator based on Bregman distances and relaxed barrier functions. Paper presented at: Proceedings of the 2019 18th European Control Conference (ECC); 2019; 1790-1795.
21. Feller C, Ebenbauer C. Relaxed logarithmic barrier function based model predictive control of linear systems. *IEEE Trans Autom Control*. 2017;62(3):1223-1238.
22. Gharbi M., Feller C., Ebenbauer C. A first step toward moving horizon state estimation based on relaxed logarithmic barrier functions. Paper presented at: Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC); 2017; 2188-2194.
23. Rockafellar RT. Monotone operators and the proximal point algorithm. *SIAM J Control Optim*. 1976;14(5):877-898.
24. Güler O. On the convergence of the proximal point algorithm for convex minimization. *SIAM J Control Optim*. 1991;29(2):403-419.
25. Parikh N, Boyd S, et al. Proximal algorithms. *Found Trends Optimiz*. 2014;1(3):127-239.
26. Censor Y, Zenios SA. Proximal minimization algorithm with D-functions. *J Optim Theory Appl*. 1992;73(3):451-464.
27. Téboulle M. Entropic proximal mappings with applications to nonlinear programming. *Math Oper Res*. 1992;17(3):670-690.
28. Lin W, Byrnes CI. Remarks on linearization of discrete-time autonomous systems and nonlinear observer design. *Syst Control Lett*. 1995;25(1):31-40.
29. Krener AJ, Isidori A. Linearization by output injection and nonlinear observers. *Syst Control Lett*. 1983;3(1):47-52.
APPENDIX A. BREGMAN DISTANCES

We give some background information on Bregman distances. Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex differentiable function. From $\psi$, we construct the Bregman distance

$$D_\psi(x, y) = \psi(x) - \psi(y) - (x - y)^T \nabla \psi(y), \quad (A1)$$

which satisfies

$$D_\psi(x, y) \geq 0 \quad \forall x, y \in \mathbb{R}^n, \quad (A2)$$

and

$$D_\psi(x, y) = 0 \iff x = y. \quad (A3)$$

Note that if we choose $\psi(z) = \frac{1}{2} ||z||^2$, we recover the squared Euclidean distance $D_\psi(x, y) = \frac{1}{2} ||x - y||^2$. Although Bregman distances are similar to distance functions, they are, in general, not symmetric and do not satisfy the triangle inequality. Nevertheless, it has been shown that Bregman distances possess the following useful property.26

**Lemma 3.** Let $\Omega \subset \mathbb{R}^n$ be a nonempty closed convex set. Let $\overline{x} \notin \Omega$ and $x^*$ denote the closest point in $\Omega$ to $\overline{x}$ with respect to the Bregman distance $D_\psi$, that is,

$$x^* = \arg \min_x \{ D_\psi(x, \overline{x}) : x \in \Omega \subset \mathbb{R}^n \}. \quad (A4)$$

Let $x \in \Omega$. Then

$$D_\psi(x^*, \overline{x}) + D_\psi(x, x^*) \leq D_\psi(x, \overline{x}). \quad (A5)$$

APPENDIX B. INPUT-TO-STATE STABILITY

We introduce some basics regarding the concept of input-to-state stability for discrete-time systems. We consider discrete-time systems of the form

$$x_{k+1} = f(x_k, w_k), \quad (B1)$$

where $x_k \in \mathbb{R}^n$ and $w_k \in \mathbb{R}^n$ refer to the state and input vectors. In words, input-to-state stability states that, for any bounded input sequence $\mathbf{w} = \{w_0, w_1, \ldots\}$, the state remains bounded with a well-defined gain between the size of the state bound and the input magnitude. This is stated more accurately in the following definition.
Definition 1. System (B1) is ISS with respect to \( w_k \) if there exists a \( \mathcal{KL} \)-function \( \beta \) and a \( \mathcal{K} \)-function \( \gamma \) such that, for each bounded input sequence \( w \), it holds that

\[
\|x_k\| \leq \beta(\|x\|, k) + \gamma(\|w_{[k-1]}\|),
\]

for any \( k \in \mathbb{N} \), where \( x_k \) refers to the solution of (B1) at time instant \( k \) with initial condition \( \xi = x_0 \).

In particular, the state of an ISS system decays for a decaying input and the origin \( x = 0 \) is globally asymptotically stable whenever \( w_k = 0 \) for all \( k \in \mathbb{N} \).

Definition 2. A continuous function \( V : \mathbb{R}^n \rightarrow \mathbb{R} \) is an ISS-Lyapunov function for (B1) if there exist \( K_\infty \)-functions \( \alpha_1, \alpha_2, \alpha_3 \), and an \( K \)-function \( \sigma \) such that

\[
\alpha_1(\|x_k\|) \leq V(x_k) \leq \alpha_2(\|x_k\|),
\]

and

\[
V(x_k) - V(x_{k-1}) \leq -\alpha_3(\|x_{k-1}\|) + \sigma(\|w_{[k-1]}\|).
\]

for any \( k \) and any inputs \( w \).

Theorem 3. If the system (B1) admits an ISS-Lyapunov function, then it is ISS.

**APPENDIX C. ISS GAINS (THEOREM 2)**

From the proof of Theorem 2, we have

\[
-\|e_{k-N-1}\|^2 \leq -\frac{1}{\beta} V(e_{k-N-1}).
\]

Defining \( \beta := 1 - c_1/\bar{p} \), we, therefore, obtain in view of (58)

\[
0 \leq V(e_{k-N}) \leq \beta V(e_{k-N-1}) + c_2 \|d_{k-N-1}\|^2.
\]

From the LMI in (50), we can see that

\[
0 < Q \preceq P - (A - LC)^TP(A - LC) \preceq P.
\]

Hence, \( c_1 < \frac{1}{\bar{\lambda}} \leq \lambda_{\text{max}}(Q) \leq \bar{p} \) and, therefore, \( \beta > 0 \). Furthermore, \( \beta < 1 \). Thus, \( \beta \in (0, 1) \). By performing a geometric series, we get

\[
V(e_{k-N}) \leq \beta^{k-N} V(e_0) + c_2 \frac{1 - \beta^{k-N}}{1 - \beta} \|d_{[0:k-N-1]}\|^2
\]

\[
\leq \beta^{k-N} V(e_0) + c_2 \frac{1}{1 - \beta} \|d_{[0:k-N-1]}\|^2.
\]

Therefore, we obtain the following inequality

\[
\frac{1}{\beta} \|e_{k-N}\|^2 \leq \beta^{k-N} \frac{1}{\bar{p}} \|e_0\|^2 + c_2 \frac{1}{1 - \beta} \|d_{[0:k-N-1]}\|^2,
\]

and hence for any \( k \geq N \)

\[
\|e_{k-N}\| \leq c_1 \beta^{k-N} \|e_0\| + c_d \|d_{[0:k-N-1]}\|,
\]

where \( \lambda_{\text{max}}(Q) \leq 1/\beta \).
where $c_e := \sqrt{p/p} \beta_e = \sqrt{\beta}$, and $c_d := \sqrt{c_2/((1 - \beta)p^2)}$. Using (54) and the triangle inequality, it holds that

$$\|d_{0:k-N-1}\| = \sup_{0 \leq l \leq k-N-1} \left\{ \|\Delta \phi_l + w_l - L v_l - P^{-1} \sum_{i=l+1}^{i+N} (CA^{i-1})^T \nabla r \left( v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \right)\| \right\}$$  \hspace{1cm} (C7)

$$\leq \sup_{0 \leq l \leq k-N-1} \left\{ \|\Delta \phi_l\| + \sup_{0 \leq l \leq k-N-1} \{\|w_l\|\} + \sup_{0 \leq l \leq k-N-1} \{\|L v_l\|\} \right\}$$  \hspace{1cm} (C8)

$$+ \sup_{0 \leq l \leq k-N-1} \left\{ \left\| P^{-1} \sum_{i=l+1}^{i+N} (CA^{i-1})^T \nabla r \left( v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \right) \right\| \right\}.$$

Since the nonlinearity $\phi$ is assumed in A2 of Assumption 3 to be Lipschitz continuous, we have

$$\sup_{0 \leq l \leq k-N-1} \{\|\Delta \phi_l\|\} = \sup_{0 \leq l \leq k-N-1} \{\|\phi(y_l) - \phi(y_{l+1})\|\} \leq L_{\phi} \sup_{0 \leq l \leq k-N-1} \{\|y_l - y_{l+1}\|\} = L_{\phi} \sup_{0 \leq l \leq k-N-1} \{\|v_l\|\}. \hspace{1cm} (C9)$$

Hence, we obtain

$$\|d_{0:k-N-1}\| \leq \|w_{0:k-N-1}\| + (\tilde{I} + L_{\phi}) \|v_{0:k-N-1}\|$$

$$+ \frac{1}{p} \sup_{0 \leq l \leq k-N-1} \left\{ \sum_{i=l+1}^{i+N} \|CA^{i-1}\| \left\| \nabla r \left( v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \right) \right\| \right\}.$$

In view of the Lipschitz continuity of $\nabla r$ given in Assumption 3, we obtain by setting $x = v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j$ and $y = 0$ in (34) that

$$\left\| \nabla r \left( v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \right) \right\| \leq \left\| v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \right\|.$$

and hence

$$\|d_{0:k-N-1}\| \leq \|w_{0:k-N-1}\| + (\tilde{I} + L_{\phi}) \|v_{0:k-N-1}\| + \frac{L_{\phi}}{p} \sup_{0 \leq l \leq k-N-1} \left\{ \sum_{i=l+1}^{i+N} \|CA^{i-1}\| \left\| v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \right\| \right\}.$$

Let $l = l^*, 0 \leq l^* \leq k - N - 1$, denote the index such that the last term in the right-hand side of the latter inequality achieves its maximum and let

$$\gamma := \sup_{0 \leq l \leq k-N-1} \left\{ \sum_{i=l+1}^{i+N} \|CA^{i-1}\||v_l + C \sum_{j=l+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \| \right\} = \sum_{i=l^*+1}^{i+N} \|CA^{i-l^*-1}\| \left\| v_l + C \sum_{j=l^*+1}^{i-1} A^{i-j-1} w_j + \Delta \phi_j \right\|.$$

Writing out the sum in the latter equality yields

$$\gamma = \|C\|\|v_{l^*+1}\| + \|CA\|\|v_{l^*+2}\| + Cw_{l^*+1} + C\Delta \phi_{l^*+1} + \ldots$$

$$+ \|CA^{N-1}\|\|v_{l^*+N}\| + CA^{N-2}w_{l^*+1} + C\Delta \phi_{l^*+1} + \ldots + Cw_{l^*+N-1} + C\Delta \phi_{l^*+N-1} \|$$

$$\leq \tilde{c} \|v_{1:k-N}\| + \tilde{c} \alpha^N \left( \|v_{1:k-N}\| + \tilde{c} L_{\phi} \|v_{1:k-N}\| + \tilde{c} L_{\phi} \|v_{1:k-N}\| + \ldots + \tilde{c} \|w_{1:k-N}\| + \tilde{c} L_{\phi} \|v_{1:k-N}\| + \ldots \right)$$

$$+ \tilde{c} \alpha^{N-1} \left( \|v_{1:k-N}\| \right) + \tilde{c} \alpha^{N-2} \left( \|v_{1:k-N}\| \right) + \tilde{c} L_{\phi} \|v_{1:k-N}\| + \ldots + \tilde{c} \|w_{1:k-N}\| + \tilde{c} \|w_{1:k-N}\| \right) \|$$

$$\leq \tilde{c} \|v_{1:k-N}\| + \tilde{c} \alpha \left( (1 + \tilde{c} L_{\phi})\|v_{1:k-N}\| + \tilde{c} \|w_{1:k-N}\| + \ldots \right) \ldots \ldots
\[
\begin{align*}
&+ \bar{c} a^{-N-1} \left( 1 + N \overline{c} L_\phi \right) \| v_{1:k-1} \| + \bar{c} a^{-N-2} \| w_{1:k-2} \| + \cdots + \bar{c} |w_{1:k-2}| \\
= & \bar{c} \sum_{i=0}^{N-1} \overline{a}^i \left( 1 + \overline{i} \overline{c} L_\phi \right) \| v_{1:k-1} \| + \bar{c}^2 \sum_{i=1}^{N-1} \overline{a}^i \sum_{j=0}^{i-1} \overline{a}^j \| w_{1:k-2} \|.
\end{align*}
\] (C14)

Substituting (C15) into (C12) yields

\[
\| \overline{d}_{0:k-N-1} \| \leq \| \overline{w}_{0:k-N-1} \| + \left( \overline{i} + L_\phi \right) \| v_{0:k-N-1} \| + \frac{L_r}{P} \bar{c} \sum_{i=0}^{N-1} \overline{a}^i \left( 1 + \overline{i} \overline{c} L_\phi \right) \| v_{1:k-1} \| + \frac{L_r}{P} \bar{c}^2 \sum_{i=1}^{N-1} \overline{a}^i \sum_{j=0}^{i-1} \overline{a}^j \| w_{1:k-2} \|
\leq \left( 1 + \frac{L_r}{P} \bar{c}^2 \sum_{i=1}^{N-1} \overline{a}^i \sum_{j=0}^{i-1} \overline{a}^j \right) \| \overline{w}_{0:k-1} \| + \left( \overline{i} + L_\phi + \frac{L_r}{P} \bar{c} \sum_{i=0}^{N-1} \overline{a}^i \left( 1 + \overline{i} \overline{c} L_\phi \right) \right) \| v_{0:k-1} \|.
\] (C15)

Substituting the latter inequality into (C6) yields the desired ISS property with respect to the disturbances.