Abstract

This paper presents a Moving Horizon Estimation (MHE) method for discrete-time nonlinear systems decomposed into coupled subsystems with non-overlapping states. In the proposed algorithm, each subsystem solves a reduced-order MHE problem to estimate its own state based on the estimates computed by its neighbors. Conditions for the convergence of the estimates are investigated. The algorithm is applied to a model of three river reaches.

Keywords: Large scale systems, moving horizon estimator, system partitioning, hydro power valley.

1. Introduction

Many industrial processes and physical systems are composed by a large number of interconnected units, such as industrial plants [21] and power networks [22]. Their control poses many algorithmic and technological problems, due for instance to their large-scale nature and to limitations in information exchange. These problems motivate the intense research activity in the design of distributed control systems, in particular with Model Predictive Control (MPC) [19]. However, most of the distributed control methods proposed so far are state-feedback, so that in order to guarantee a fully dis-
tributed control design, also distributed state estimation algorithms dealing with constraints are needed. Early works in distributed estimation were aimed at reducing the computational complexity of centralized Kalman filters by parallelizing computations, see e.g. [8, 15], under the assumption that each subsystem has full knowledge of the whole dynamics. Subsequently, [13] proposed a solution based on the use of reduced-order and decoupled models for each subsystem, while subsystems with overlapping states were considered in [10, 20, 21]. More recently, consensus-based distributed state estimators for sensor networks where each sensor measures just some of the system outputs and computes the estimate of the overall state have been studied, e.g. see [14]. Along the same lines and in order to cope with constraint on noise and state variables, in [4, 5] distributed Moving Horizon Estimators (MHE) for sensor networks have been proposed. Partition-based MHE algorithms (PMHE) for linear constrained systems decomposed into interconnected subsystems without overlapping states have recently been developed in [6]. In these algorithms, each subsystem estimates its own states based on information transmitted by its neighbors.

The aim of this paper is to extend the results of [6] to the case of nonlinear systems. The convergence properties of the method are guaranteed under suitable sufficient conditions.

The proposed partition-based MHE is applied to the problem of estimating the levels and flow rates in a model of three cascade river reaches. Interconnections between successive reaches are due to the dependence of the input flow rate of the downstream reaches to the level of the final section of the upstream ones, which cannot be measured, but just estimated from the available measures collected along the reach.

The paper is structured as follows. Section 2 introduces nonlinear partitioned systems and the main assumptions concerning their dynamics. Section 3 describes the proposed MHE algorithm, while convergence results are provided in Section 4. The illustrative example is considered in Section 5. For the sake of readability, the proofs of the main results are collected in an Appendix.

**Notation.** $I_n$ and $0$ denote the $n \times n$ identity matrix and the matrix of zero elements whose dimensions will be clear from the context. The notation $\|z\|_S^2$ stands for $z^T S z$. 

2
where $S$ is a symmetric positive-semidefinite matrix. Given a set of scalar variables $v = \{v^1, \ldots, v^m\}$, $i_1 < i_2 < \cdots < i_n$, we use the short-hand $(v)$ or $(v^i_1, \ldots, v^i_m)$ to denote the vector $[v^i_1, \ldots, v^i_m]^T$. By extension, if $v^i_1, \ldots, v^i_m$ are sets of scalar variables, $(v)$ or $(v^i_1, \ldots, v^i_m)$ denote $((v^i_1), \ldots, (v^i_m))$. With a little abuse of notation, sometimes we will use $v$ instead of $(v)$. Given a function $f(x_1, \ldots, x_n): D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ we define $\text{arg}(f) = \{x_i : f \text{ is not constant in } x_i \text{ on } D\}$.

For a discrete-time signal $w_t$ and $a, b \in \mathbb{N}$, $a \leq b$, $w_{[a:b]} = (w_a, w_{a+1}, \ldots, w_b)$. For the definition of positive-definite, $K$, $K\infty$ and $K_L$ functions we defer the reader to [17]. Finally, given $x_k, x_k^* \in \mathbb{R}^n$, we define $\|x_k - x_k^*\|_{[a:b]} = \max_{k \in [a:b]} \|x_k - x_k^*\|$, where $\| \cdot \|$ denotes the Euclidean norm.

**2. Nonlinear large-scale systems**

We consider the discrete-time nonlinear system

$$x_{t+1} = f(x_t, u_t) + w_t,$$

(1)

where $x_t = (x^1_t, \ldots, x^n_t) \in \mathbb{R}^n$ is the state, $w_t \in \mathbb{R}^n$ is the process noise and $u_t \in \mathbb{R}^m$ is the input. Measurements are performed according to the sensing model

$$y_t = h(x_t, u_t) + v_t$$

(2)

where $v_t \in \mathbb{R}^p$ is the measurement noise. We assume that $f(x, u, t)$ has continuous partial derivatives with respect to the argument $x$ and satisfies the following Assumption.

**Assumption 1.** Function $f: \mathcal{X} \times \mathbb{R}^m \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$, where $\mathcal{X} \subseteq \mathbb{R}^n$, is globally Lipschitz with respect to $x$ i.e., $\exists l_x > 0 : \forall x_1, x_2, u \in \mathcal{X}$ and

$$\|f(x_1, u) - f(x_2, u)\| \leq l_x \|x_1 - x_2\| \quad (3)$$

We split system (1) into $M$ interconnected submodels by choosing a time-invariant partition of elements of $x_t$ into the sets $x^i_t, i = 1, \ldots, M$, so that $x_t = (x^1_t, \ldots, x^M_t)$ up to an index permutation, where $x^i_t \in \mathbb{R}^n$ for all $i = 1, \ldots, M$. Accordingly, we define
As for the outputs of subsystems, we assume that the sets \( y^i \), \( i = 1, \ldots, M \), are time-invariant partition of variables in \( y \) and, analogously, the sets \( v^i \), \( i = 1, \ldots, M \), are time-invariant partition of variables in \( v \), verifying

\[
y^i = h^i(x^i, u^i) + v^i
\]

for suitable functions \( h^i \). Note that (7) assumes that \( y^i \) only depends upon the local variables \( x^i \) and \( u^i \). We highlight that this structural assumption is made only for the sake of simplicity and the main results can be generalized to the case where it does not hold.

From now on, we assume that the system partitioning and the input sequence \( \{u_i\} \) are such that the following observability assumption on the local subsystems is satisfied (see [17] for \( M = 1 \)).
Assumption 2. For all \( i = 1, \ldots, M \) and \( j \in \mathcal{N}_i \), there exists \( n_i^o \in \mathbb{N} \setminus \{0\} \) and functions \( \gamma_i^o(\cdot), \gamma_j^o(\cdot) \in \mathcal{X} \) such that for every initial states \( x_i^o, x_0^o \), for any feasible sequences \( x_i^o_{[0:k-1]}, x_0^o_{[0:k-1]} \) and for any disturbance sequences \( w_i^o_{[0:k-1]}, w_0^o_{[0:k-1]} \)

\[
\| x_i^o_{0:k-1} - x_0^o_{0:k-1} \| \leq \gamma_i^o (\| w_i^o_{0:k-1} - w_0^o_{0:k-1} \|)
\]

\[
+ \gamma_j^o (\| y_j^o - y_i^o \|) + \sum_{j \in \mathcal{N}_i} \gamma_j^o (\| x_j^o - x_i^o \|)
\]

where \( k \geq n_i^o \), and \( y_i^o \) and \( y_j^o \) are the output sequences stemming from \( (w_i^o, u_i^o, x_i^o) \) and \( (w_i^o, u_i^o, x_i^o, u_i^o) \) with initial conditions \( x_i^o_{0:k} \) and \( x_0^o_{0:k} \), respectively.

3. A non-iterative moving horizon partition-based algorithm

Our aim is to design, for each subsystem, a non-iterative distributed estimation scheme based on neighbor-to-neighbor communication for computing a reliable estimate \( \hat{x}_i^o \) of \( x_i^o \) based on the measurements \( y_i^o \) and on the crosstalk terms \( u_i^o, x_i^o \) provided by the estimators associated to the other subsystems. We extend to nonlinear systems the partitioned moving horizon estimation scheme of [6], which is here denoted NPMHE.

3.1. Model for estimation and information transmission graph

We denote with \( \hat{x}_i^o t_{i/2} \) the estimate of \( x_i^o t_{i/2} \) performed at time \( t_2 \) by subsystem \( i \) and we define \( \hat{X}_{i/2} = (\hat{x}_i^o t_{i/2}, \ldots, \hat{x}_i^o M_{i/2}) \). At each instant \( t \geq N \) we assume that an estimate of the crosstalk term \( u_i^o, x_i^o \) \( k = t - N, \ldots, t \) is provided by the neighbors at time \( t - 1 \), therefore allowing for decentralization of the state estimation algorithm. At time \( t \) the estimation model is, for \( k = t - N, \ldots, t - 1 \)

\[
\hat{x}_i^o t_{k+1} = f_i^o (\hat{x}_i^o t_k, \hat{u}_i^o t_{k-1}, u_i^o t_k) + \hat{w}_k^o \quad (9a)
\]

\[
y_i^o t_k = h_i^o (\hat{x}_i^o t_k, u_i^o t_k) + \hat{v}_k^o \quad (9b)
\]

In (9), \( \hat{u}_i^o t_{k-1} \) is the set of variables \( \{ \hat{x}_i^o t_{k-1}, j \in \mathcal{N}_i \} \) i.e., the estimates of the system’s states to be transmitted to subsystem \( i \) from neighbors at time \( t \). Therefore, subsystems communicate over a network that has the same topology of the interconnection graph. Finally, note that the noise estimates \( \hat{w}_k^o \) and \( \hat{v}_k^o \) in (9) encompass both the noise appearing in the equations (5), (7) and the estimation error due to the uncertainty on \( u_i^o, x_i^o \).
3.2. The NPMHE estimation problems

Given an estimation horizon $N \geq 1$, in order to perform the NPMHE algorithm, each node $i \in \mathcal{V}$ at time $t$ solves the constrained minimization problem NMHE-$i$ defined as

$$\Theta_i^{[i]} = \min_{\hat{x}_N^{[i]}, \hat{w}^{[i]}} J^{[i]}(t - N, t, \hat{x}_N^{[i]}, \hat{w}^{[i]}, v^{[i]}, \Gamma_t^{[i]})$$  \hspace{1cm} (10)

where $\hat{w}^{[i]}$ and $v^{[i]}$ stand for $\hat{w}_{[t-N]}^{[i]}$ and $v_{[t-N]}^{[i]}$, respectively, under the constraints imposed by system (9) and

$$\hat{x}_k^{[i]} \in \mathcal{X}_i, k = t - N, \ldots, t$$ \hspace{1cm} (11)

The local cost function $J^{[i]}$ is given by

$$J^{[i]}(t - N, t, \hat{x}_N^{[i]}, \hat{w}^{[i]}, v^{[i]}, \Gamma_t^{[i]}) = \sum_{k=t-N}^T L^{[i]}(\hat{w}_k^{[i]}, v_k^{[i]}) + \Gamma^{[i]}(\hat{x}_{t-N}^{[i]}, \hat{x}_{t-N/1}^{[i]})$$ \hspace{1cm} (12)

In (12), the functions $L^{[i]}$ and $\Gamma^{[i]}_{t-N}$ are the stage cost and the initial penalty, respectively, and must be defined in order to satisfy the following assumption.

**Assumption 3.** $L^{[i]}_0$, $\Gamma^{[i]}_{t-N}$ and $\Theta^{[i]}_{t-1}$, $t > N$ are continuous, bounded, positive definite and satisfy the following inequalities for all $w^{[i]} \in \mathbb{R}^n$, $v^{[i]} \in \mathbb{R}^p$, $x_1^{[i]}$, $x_2^{[i]} \in \mathbb{R}^n$

$$\gamma_L(||(w^{[i]}, v^{[i]})||) \leq L^{[i]}(w^{[i]}, v^{[i]})$$ \hspace{1cm} (13a)

$$\gamma_0(||x_1^{[i]} - x_2^{[i]}||) \leq \gamma_0(||x_1^{[i]} - x_2^{[i]}||)$$ \hspace{1cm} (13b)

where $\gamma_L$ and $\gamma_0$ are suitable $\mathcal{X}_o$ functions.

The quantities $\hat{x}_{t-N/1}^{[i]}$ and $\left\{\hat{w}_k^{[i]}\right\}_{k=t-N}$ are the optimizers to (10) and $\hat{x}_k^{[i]}$, $k = t - N + 1, \ldots, t$ is the local state sequence stemming from $\hat{x}_N^{[i]}$, $\left\{\hat{a}_{k/1}^{[i]}\right\}_{k=t-N}$ and $\left\{\hat{w}_k^{[i]}\right\}_{k=t-N}$. Note that subsystems, during each sampling interval, solve in parallel the local optimization problems (10) and exchange information with their neighbors.

Research on efficient numerical methods for the solution of optimization problems similar to (10) is currently very active (see [3] and references therein). An algorithm with low computational burden is proposed in [23], where it is shown that very effective results can be obtained even for large-scale applications.
3.3. The collective minimization problem

We denote by $J$ the sum of the local cost functions $J_i^j$, given by (12), i.e.

$$J = \sum_{i=1}^{M} J_i^j(t-N,i,x_{i-N},p_{i-N},\Gamma_{i-N})$$

(14)

We define the collective vectors $\hat{x}_t = (\hat{x}_t^{[1]},\ldots,\hat{x}_t^{[M]}) \in \mathbb{R}^n$, $\hat{v}_t = (\hat{v}_t^{[1]},\ldots,\hat{v}_t^{[M]}) \in \mathbb{R}^p$, $\hat{w}_t = (\hat{w}_t^{[1]},\ldots,\hat{w}_t^{[M]}) \in \mathbb{R}^n$ and rewrite $J$ as

$$J = \sum_{k=t-N}^{t} L(\hat{w}_k,\hat{v}_k) + \Gamma_{t-N}(\hat{x}_{t-N};\hat{x}_{t-N/-1})$$

(15)

where $L$ and $\Gamma_{t-N}$ are given by

$$L(\hat{w}_k,\hat{v}_k) = \sum_{i=1}^{M} L_i^j(\hat{w}_k^{[i]},\hat{v}_k^{[i]})$$

(16a)

$$\Gamma_{t-N}(\hat{x}_{t-N};\hat{x}_{t-N/-1}) = \sum_{i=1}^{M} \Gamma_i^j(\hat{x}_t^{[i]};\hat{x}_{t-N/-1})$$

(16b)

We define the function

$$\tilde{f}(\hat{x}_k,\hat{x}_{k/-1},u_k) = (f^{[1]}(\hat{x}_k^{[1]},\hat{x}_{k/-1}^{[1]},u_k^{[1]}),\ldots,f^{[M]}(\hat{x}_k^{[M]},\hat{x}_{k/-1}^{[M]},u_k^{[M]}))$$

so that constraints (9) and (11) can be written in the collective form

$$\hat{x}_{k+1} = \tilde{f}(\hat{x}_k,\hat{x}_{k/-1},u_k) + \hat{w}_k$$

(17a)

$$y_k = h(\hat{x}_k,u_k) + \hat{v}_k$$

(17b)

$$\hat{x}_k \in \mathbb{X}$$

(17c)

with $k = t-N,\ldots,t$. The solution to

$$\min_{\hat{x},\ldots,\hat{w},\hat{v},\Gamma_{t-N}} J(t-N,t,\hat{x}_{t-N},\hat{w},\hat{v},\Gamma_{t-N})$$

(18)

where $\hat{w}$ and $\hat{v}$ are short-hand for $\hat{w}_{t-N/-t}$ and $\hat{v}_{t-N/-t}$ respectively, is equivalent to the solution to the MHE-$i$ problems (10), in the sense that $\hat{x}_{t-N/-t},\hat{w}_{t-N/-t}$ is a solution to (10) if and only if $\hat{x}_{t-N/-t},\hat{w}_{t-N/-t}$ is a solution to (18), where $\hat{w}_{t-N/-t} = (\hat{w}_t^{[1]},\ldots,\hat{w}_t^{[M]})$. In fact, at time $t$, variables $\hat{x}_{t-N/-t}$ are fixed inputs for the system (17).
We define the transit cost for subsystem $i$ as

$$
\Xi^{[i]}_{[t-N+1:t]}(z^{[i]}_{[t-N+1:t]}) = \min_{\tilde{x}^{[i]}_N, w} \left\{ J^{[i]}(t-N,t,\tilde{x}^{[i]}_N,\tilde{w},\tilde{v},\Gamma^{[i]}_{t-N}) \right\}
$$

subject to (9), (11) and $z^{[i]}_k = z^{[i]}_k$ for $k = t-N+1,\ldots,t$ 

and define the collective transit cost in a generic sequence $z_k = (z^{[1]}_k,\ldots,z^{[M]}_k) \in \mathbb{R}^n$ $k = t-N+1,\ldots,t$ as

$$
\Xi^{[i]}_{[t-N+1:t]}(z^{[i]}_{[t-N+1:t]}) = \min_{\tilde{x}_N, w} \left\{ J(t-N,t,\tilde{x}_N,\tilde{w},\tilde{v},\Gamma_{t-N}) \right\}
$$

subject to (17) and $\tilde{x}_k = z_k$ for $k = t-N+1,\ldots,t$  

It holds that

$$
\Xi^{[i]}_{[t-N+1:t]}(z^{[i]}_{[t-N+1:t]}) = \sum_{i=1}^{M} \Xi^{[i]}_{[t-N+1:t]}(z^{[i]}_{[t-N+1:t]})
$$

From (16) and in view of Assumption 3 there exist suitable $\mathcal{K}_{\infty}$ functions $\gamma^0_i$ and $\gamma^0_0$ such that

$$
\gamma^0_i(||(w,v)||) \leq L(w,v)  
$$

$$
\gamma^0_0(x_1;x_2) \leq \gamma^0_0(||x_1-x_2||)
$$

for all $w \in \mathbb{R}^n$, $v \in \mathbb{R}^p$ and $x_1, x_2 \in \mathbb{R}^n$.

We discuss now the choice of the initial penalties $\Gamma_{t-N}^{[i]}$.

**Assumption 4.** Given a state sequence $z_k \in \Xi$, $k = t-N,\ldots,t-1$, the following inequalities are verified

$$
\Theta^*_{t-1} \leq \Gamma_{t-N}(z_{t-N};x_{t-N/2-1})
$$

$$
\sum_{k=1}^{t-N} L\left( \tilde{f}(z_k, \tilde{x}_k, u_k) - \tilde{f}(z_k, \tilde{x}_{k/2-1}, u_k), 0 \right) + \Gamma_{t-N}(z_{t-N};x_{t-N/2-1}) \leq \Xi_{[t-N/2-1]/t-1}(z_{t-N/2-1})
$$

where $\Theta^*_{t-1} = \sum_{i=1}^{M} \Theta^{[i]}_{t-1}$. 

8
Remark 3.1. Assumption 4 is critical. In fact, even in case $M = 1$ (i.e., trivial partition) where one has that $\sum_{k=t-N}^{t-1} L(\hat{f}(z_k, \hat{x}_k, u_k) - \hat{f}(z_k, \hat{x}_{k/t-1}, u_k), 0) = 0$, there are just approximate methods for computing initial penalties $\Gamma_{t-N} \neq \Theta_{t-1}$ verifying (23) and (24), see [17].

However, $z_{t-N} = \hat{x}_{[t-N]} / t-1$ are minimizers of $\Xi_{[t-N]} / t-1(z_{t-N})$ and yield $\Xi_{[t-N]} / t-1(z_{t-N}) = \Theta_{t-1}$. Hence (23)-(24) imply that $\hat{x}_{[t-N]} / t-1$ is a minimizer of $\Gamma_{t-N}$ and that $\Gamma_{t-N}(\hat{x}_{t-N} / t-1; \hat{x}_{t-N} / t-1) = \Theta_{t-1}$.

As in [6], a choice for $\tilde{L}[i]$ and $\tilde{\Gamma}[i]_{t-N}$ fulfilling Assumption 3 and (23) is

$$L[i] = \frac{1}{2} ||\tilde{v}[i]_k ||^2_{(\tilde{Q}[i])^{-1}} + \frac{1}{2} ||\tilde{v}[i]_k ||^2_{(\tilde{R}[i])^{-1}},$$

$$\Gamma[i]_{t-N} = \frac{1}{2} ||\tilde{x}[i]_{t-N} - \hat{x}[i]_{t-N} / t-1 ||^2_{(\Pi[i]_{t-N})^{-1}}, + \tilde{\Theta}[i]_{t-1}$$

where $\tilde{Q}[i]$ and $\tilde{R}[i]$ are suitable symmetric and positive-definite matrices, and $\Pi[i]_{t-N}$ is a symmetric semi definite-positive matrix.

It is easy to prove that, under Assumption 1 and if the stage cost and the initial penalty are defined as in (25) and (26), respectively, Assumption 4 is verified if, for each subsystem $i$, the following inequality is satisfied, for all sequences $z[i]_{t-N} \in \mathcal{X}_i$, $k = t - N, \ldots, t - 1$

$$\frac{1}{2} \sum_{k=t-N}^{t-1} q[i] ||z[i]_k - \tilde{x}[i]_{k/t-1} ||^2 + \frac{1}{2} ||\tilde{x}[i]_{t-N} - \hat{x}[i]_{t-N} / t-1 ||^2_{(\Pi[i]_{t-N})^{-1}}, + \tilde{\Theta}[i]_{t-1} \leq \Xi_{t-N} / t-1(z_{t-N})$$

where, for all $i$, $q[i] = \sum_{j=1}^{M} I_{ij} I_{ij} = M I_{ij} \sigma_{\max} \left( (\tilde{Q}[i])^{-1} \right)$ if $j \in \mathcal{N}_i$ and $I_{ij} = 0$ otherwise, and $\sigma_{\max}(\cdot)$ denotes the maximum singular value of its argument.

Note that (27), in view of this, the implementation of the NPMHE estimation scheme results to be completely decentralized.

4. Convergence properties of the proposed estimators

In this section the convergence results reported in [16] for centralized estimators are extended to the proposed NPMHE scheme. Similarly to [16], these properties are analyzed in a deterministic setting.
Figure 1: River scheme.

**Definition 1.** Let $\Sigma$ be system (1) with $w_t = 0$ and denote by $x_{\Sigma}(t, x_0)$ the state reached by $\Sigma$ at time $t$ starting from initial condition $x_0$, $k = 0, \ldots, t - 1$. Assume that the trajectory $x_{\Sigma}(t, x_0)$ is feasible, i.e., $x_{\Sigma}(t, x_0) \in \mathcal{X}$ for all $t$. Then, NPMHE is asymptotically convergent if $\|\hat{x}_{ij}(t) - x_{\Sigma}(t, x_0)\| \to 0$.

Moreover, in order to state the main convergence result of the paper, some further definitions are required.

**Definition 2.** Let $\eta \geq 0$ and $r$ be a real and an integer number respectively. Then, $\lambda(\eta, r)$ is a function defined as

$$\lambda(\eta, r) = \frac{1 - \eta^r}{1 - \eta}, \text{ if } \eta \neq 1, \text{ and } \lambda(\eta, r) = r, \text{ if } \eta = 1$$

**Definition 3.** The $\mathcal{X}_\eta$-functions $\mathcal{H}_j$ are defined as

$$
\mathcal{H}_j(\eta) = \max_{k \in [r-N+1]} \left( \frac{1}{\eta^{k-t-N}} \mathcal{H}_j(\eta) + i_j \lambda(\eta, k - (t-N)) \right) \quad (28)
$$

**Definition 4.** (see [2]). Given a vector $\delta \in \mathbb{R}^M$, with scalar components $\delta[i] \geq 0$, $i = 1, \ldots, M$, we define the gain map

$$
\Gamma(\delta) = \begin{bmatrix} 
\sum_{j \in \mathcal{K}_1} \mathcal{H}_j(\delta[i]) \\
\vdots \\
\sum_{j \in \mathcal{K}_M} \mathcal{H}_M(\delta[i]) 
\end{bmatrix} \quad (29)
$$

and the diagonal operator $D : \mathbb{R}^M \to \mathbb{R}^M$ such that

$$
D(\delta) := \begin{bmatrix} 
(Id + d_1)(\delta[1]) \\
\vdots \\
(Id + d_M)(\delta[M]) 
\end{bmatrix}
$$
where \( \text{Id} \) is the identity function and \( d_i \in \mathcal{K} \), \( i = 1, \ldots, M \).

Finally we need to introduce a small gain condition, which will be fundamental to guarantee convergence of NPMHE and is derived from [2]. Specifically, we will require that, for all \( \delta \neq 0 \), with components \( \delta^{[i]} \geq 0, i = 1, \ldots, M \) one has

\[
\tilde{\Gamma} \circ D \not\geq \text{Id} \quad (30)
\]

Inequality (30) requires that, for all \( \delta \), at least one component of vector \( \tilde{\Gamma}(D(\delta)) \) strictly decreases. As shown in the proof of Theorem 1, the quantity \( \delta^{[i]} \) is related to the magnitude of the estimation error for the subsystem \( i \) over a time horizon, and \( \tilde{\Gamma} \) models the mutual interactions of \( \delta^{[i]} \) variables. Furthermore, in the linear framework, \( \tilde{\Gamma} \) is a square matrix of system gains and condition (30) amounts to verify that \( \tilde{\Gamma} \) is Schur. For further details we defer the reader to [2].

**Theorem 1.** If Assumptions 1, 2, 3 and 4 hold and if, for all \( \delta \neq 0 \), with components \( \delta^{[i]} \geq 0, i = 1, \ldots, M \), (30) is verified, then NPMHE is asymptotically convergent.

Recall that a Directed Acyclic Graph (DAG) is a directed graph with no cycles. Namely, \( \mathcal{G} \) is a DAG if, for all subsystems \( i \) and \( j \), when there is a path from \( i \) to \( j \), then there does not exist a path from \( j \) to \( i \).

**Corollary 1.** If Assumptions 1, 2, 3 and 4 hold and if the interconnection graph is a DAG, then the NPMHE scheme is asymptotically convergent.

### 5. An application: three cascade river reaches

In this section we apply NPMHE to a system composed by three river reaches. The development of distributed predictive control techniques for this system requires proper estimation of the states of the reaches, i.e. levels and flow rates at different points, as well as of the disturbances, which correspond to unknown input/output terms (e.g., creeks, rain, leakages).
5.1. Model of the reaches

Each reach is endowed with a power house placed in a lateral channel where a turbine generates the electric power, and with a weir along the main natural river course, see Fig.1. The model of a single reach is based on the classical de Saint Venant equations, see e.g. [18], [7], [11], [12]. Letting $x \in \mathbb{R}$ be the main spatial coordinate defined by the flow direction and $\tau$ be the continuous time, the assumptions of constant river width $W(x)$ and rectangular cross section $S(x, \tau)$ are made. Furthermore, we assume that the river friction slope is given by the Manning-Strickler equation [11]. According to the previous assumptions and denoting with $H(x, \tau)$ and $Q(x, \tau)$ the river height and the flow rate, respectively, we write the de Saint Venant equations as

$$\frac{\partial H}{\partial \tau} = -\frac{1}{W} \frac{\partial Q}{\partial x}$$
$$\frac{\partial Q}{\partial \tau} = 2Q \frac{\partial H}{\partial x} + \left( \frac{Q}{H} \right)^2 \left( \frac{1}{W} - gWH \right) \frac{\partial H}{\partial x} + gWL_0H$$

where, for simplicity, we omit the dependence of the variables $Q$ and $H$ upon $x$ and $\tau$, $g$ is the gravitational acceleration, $L_0$ is the bed slope and $k_{str}$ is the Strickler coefficient.

For simulation and estimation purposes, we discretize the model into $N_c$ sections along the flow direction, each one with length $\Delta x = X/N_c$, where $X$ is the total length of the reach. To avoid unnecessary stiffness, the crossing sections of the different variables are overlapped. The flow rates $Q$ are computed at the crossing of each section while the heights $H$ are computed in the middle of the section, see Fig.2 and the references [18], [7]. The discretization is made by the finite difference method by approximating the derivatives with the first term in the Taylor series expansions of $Q$ and $H$ around any point $\chi_i$:

$$\frac{\partial Q_2}{\partial \chi_2} = \frac{Q(\chi_{2i+1}) - Q(\chi_{2i})}{\chi_{2i+2} - \chi_{2i}}$$
$$\frac{\partial H_{2i-1}}{\partial \chi_{2i-1}} = \frac{H(\chi_{2i+1}) - H(\chi_{2i-1})}{\chi_{2i+1} - \chi_{2i-1}}$$

The boundary conditions are given by the inlet flow rate at the first reach $Q_{in}$ and by the output flow rate $Q_{out}$, which is a function of the level at the end of the third reach, i.e. $Q_{out} = Q_{out}(H_{2N_c+1})$. For the second and third reaches, the inlet flow rates coincide with the outlet flow rate of the upstream reach. The output flow rate of each reach is

$$Q_{out} = Q_l + Q_D(H_{2N_c+1})$$

12
where

\[ Q_D(H_{2N_c+1}) = k_{weir} A_{weir} \sqrt{2g(H_{2N_c+1})} \]  \hspace{1cm} (33)

\( k_{weir} \) is a parameter that depends on the characteristic of the dam, \( A_{weir} \) is the weir cross-sectional area, \( Q_D \) is the flow rate through the weir of the dam and \( Q_t \) is the flow rate through the channel and the power house, assumed constant in the considered state estimation problem. In the following we divide each reach into \( N_c = 5 \) cells.

5.2. Disturbances model

We assume that each reach is affected by an unmeasurable inlet flow rate \( \omega_{l1} \) acting at the beginning of each reach and which represents the variation of the inflow \( Q_{in} \) due to the variations of the concession level of an upstream dam. Moreover, an additional flow rate \( \omega_{l2} \) is placed at the input of the third cell of the second reach to model the presence of an unknown affluent. Both \( \omega_{l1} \) and \( \omega_{l2} \) are generated as the sum of a constant term (\( \omega_{1} \) and \( \omega_{2} \)) and the state (\( d_{1} \) and \( d_{2} \)) of a first-order stable system fed by zero-mean White Gaussian Noises (WGN) \( w_{1}(k) \) and \( w_{2}(k) \) with variances \( \sigma_{1}^{2} \) and \( \sigma_{2}^{2} \), respectively. We include saturation constraints to impose that these disturbances are non negative.
5.3. River data and available measurements

The three reaches have the same geometric characteristics: the length and the width of each reach are 4000[m] and 100[m], respectively, the slope of the bed of each reach \( I_0 \) is 0.0033, the section \( A_{weir} \) and the discharge coefficient \( k_{weir} \) of the weir of the dam are 18.26[m²] and 0.6, respectively, the nominal flow rate through the turbine \( Q_t \) is 100[m³/s], and the Strickler coefficient \( k_{str} \) is 30[m¹/³/s]. In nominal stationary conditions, the considered constant flow rate is \( \bar{Q} = 300[m³/s] \), while the values of the height are \( \bar{H}_1 = 3.83[m], \bar{H}_3 = 7.11[m], \bar{H}_5 = 10.4[m], \bar{H}_7 = 13.7[m], \bar{H}_9 = 17[m] \). As for the disturbances, we use the following values: \( \bar{\omega}_1 = 10[m³/s], \bar{\omega}_2 = 30[m³/s], \sigma_1^2 = \sigma_2^2 = 5 \), while the filters have gain equal to 0.5 and time constant 10³[s]. Moreover, we assume that the inflow \( Q_m \) of the first reach is known, as well as the flow rates through the turbines, while for estimation and control purposes three measurements are available for any reach, namely the levels \( H_1 \) and \( H_5 \) and the flow rate \( Q_8 \) in the first and third reaches, and the variables \( H_1 \), \( H_7 \) and \( Q_4 \) in the second reach. These measurements are affected by noise; a WGN with zero mean and variance \( \sigma^2 = 0.1 \) is added to the level measures, while a zero mean WGN with unitary variance corrupts the flow measure. Remarkably, the measured flow \( Q_8 \) does not correspond to the inflow of the downstream reach, see eqs. \( (32) \) and \( (33) \), so that there is an effective coupling between the estimation problems. This motivates the use of the NPMHE scheme presented in the previous sections.

5.4. MHE and simulation results

We apply NPMHE to the three reaches viewed as a system with a cascade structure; therefore, Corollary 1 is automatically verified. The reaches are described by the equations arising from the spatial discretization of 31, while the mutual influences are due to the relations \( (32) \) and \( (33) \). As such, for \( i = 1, 2, 3 \), the state of the subsystems are \( x[i] = (H_1[i], Q_2[i], H_3[i], Q_4[i], H_5[i], Q_6[i], H_7[i], Q_8[i], H_9[i]) \), while the inputs are \( u[i] = Q_t \). Positivity constraints on all the estimated states have been imposed and flows \( Q_j[i] \) (\( i = 1, 2, 3, j = 2, 4, 6, 8 \)) are constrained to be smaller than 450[m³/s]. Furthermore, we consider additional constraints for each reach, on the absolute value of the difference between subsequent fluxes and levels e.g., \( |Q_2[j] - Q_2[j+2]| \leq 30 \text{ m³/s}, |Q_2[j] - Q_2[j+2]| \leq 150 \text{ m³/s} \).
m³/s, and |Q^3_{2j} - Q^3_{2j+2}| ≤ 65 m³/s, j = 1, 2, 3 and |H^i_{2(j−1)+1} - H^i_{2j+1}| ≤ 5 m, for all i = 1, 2, 3, and j = 1, 2, 3.

We discretize the models of the reaches with a sampling time Δτ = 60[s] to implement NPMHE that assumes discrete-time systems. In the discrete-time model so obtained, we assume that the state disturbance w_t (see (1)) acts on the states H_1, Q_2 for the first and third reaches, and on H_1, Q_2, H_5, Q_6 for the second reach.

The stage cost and the initial penalty (see (12)) are given by equations (25) and (26), where \( \Pi^i_{t-N/\mu-1} = 10 I_9 \), i = 1, 2, 3. The matrices \( Q^i \) are diagonal with elements equal to \( 3.33 \cdot 10^4 \) for the non-zero components of \( w_t \), and equal to very small values for its zero components. Also the matrices \( R^i \) have been chosen as diagonal, with elements equal to 200 for the level estimation errors and to \( 2 \cdot 10^3 \) for the flow rate estimation errors.

We perform the simulation experiments with MATLAB and we carry out optimizations with the TOMLAB optimization environment [9]. We add a sinusoidal variation of amplitude ±30[m³/s] and period of about 2.3[h] to the nominal inlet flow rate \( Q_{in} \).

In Figs. 3-5 we show the results obtained by applying NPMHE to a case where the estimator initial conditions are in the range of 10% with respect to the real system’s initial conditions. Figs.3 and 4 show the true and estimated values of the levels \( H^i_9 \) and of the flow rates \( Q^i_2 \), respectively, for i = 1, 2, 3. Finally, Fig.5 depicts the true and estimated disturbances both acting on the initial section of the three reaches and on the second section of reach 2. These results clearly show that, after an initial transient partly due to the optimization procedure, all the estimates converge to the true values. We also show a comparison with a centralized MHE (cMHE) algorithm [16] in terms both of estimation accuracy and of computational burden required to perform the optimization task. The estimation accuracy is evaluated in terms of the mean of the square estimation error \(|x_t - \hat{x}_t|^2\), computed both in stationary conditions \( t \in [2400, 16000] \) s and including the initial transient \( t \in [1, 16000] \) s, denoted sMSE and nsMSE, respectively. With cMHE, we obtain sMSE=15.19 and nsMSE=157.16 while, with NDMHE, sMSE=15.53 and nsMSE=196.47. This results show that NPMHE produces a small degradation in estimation accuracy with respect to the centralized solution, mainly
during the initial transient. The average computational time required to solve cMHE is $T_c = 138.27$ s while optimization problems NMHE-$i$, for reaches 1, 2 and 3, require
$T_1 = 8.21 \text{ s}$, $T_2 = 29.64 \text{ s}$ and $T_3 = 44 \text{ s}$, respectively. Therefore, in the considered scenario, the computational burden required for solving cMHE can become prohibitive if the number of subsystems $M$ is large, while the computational load required to solve NPMHE for each single subsystem is independent of $M$.

Finally, we test the reliability of NPMHE with respect to different estimator initial conditions generated as follows. We compute, for each reach ($i = 1, 2, 3$), the steady states corresponding to flux boundary conditions $Q_i^{(j)}$ of 150, 225, 375 and 450 m$^3$/s and level boundary conditions $H_i^{(j)}$ of 8, 12.5, 21.5 and 26 m. Then we set initial estimates to these steady states. This procedure allowed us to generate 16 physically meaningful initial estimates corresponding to variations of fluxes and levels spanning more than 50% of their nominal values. NPMHE performances are evaluated using the variables $H_{i}^{(2)}$ and $Q_{i}^{(2)}$, that correspond to the level and the flux displaying higher estimation errors. In the sequel, $e_{i,j} = x_i - \hat{x}_{i,j}$ denote the estimation error for a generic scalar variable $\hat{x}_{i,j}$ and is its estimate obtained in experiment $j = 1, \ldots, 16$. Fig.6 shows, for both $H_{i}^{(2)}$ and $Q_{i}^{(2)}$, the evolution of the mean estimation error $\bar{e}_t = (1/16) \sum_{j=1}^{16} e_{j,t}$ and $\bar{e}_t \pm 2\sigma$, being $\sigma$ the standard deviation of $e_{j,t}$ i.e., $\sigma^2 = (1/16) \sum_{j=1}^{16} (e_{j,t} - \bar{e}_t)^2$. It is apparent that, while wrong initial priors affect the estimation performance during the
initial transient, in stationary conditions this effect asymptotically vanishes.

![Graph 1](image1)

**Figure 6**: Mean estimation error of $x_t$ (solid line) ± $\sigma_t$ (dashed lines), for $H_2^2$ (above) and $Q_6^2$ (below).

6. Discussion and conclusions

A novel distributed moving horizon estimation scheme has been proposed for large scale systems with nonlinear dynamics which can be partitioned into a number of sub-systems with non overlapping states. The convergence properties of the method have been analyzed and sufficient conditions have been derived. The proposed approach has been applied to three river reaches and simulation results confirm the viability of the proposed approach.

Future work will be devoted to the derivation of weaker convergence conditions as well as to the development of automatic methods for the tuning of the algorithm’s free design parameters, i.e. the disturbance and measurement errors weights in the performance index. This can be of great importance, not only to guarantee convergence, but also to reduce the effect of disturbances on the estimation error. In the deterministic centralized setting, the effect of bounded disturbances on the estimator quality has been studied in [1]. Finally, analyses of the potentialities of different design choices will be...
carried out. For instance, the possibility of exploiting asynchronous transmission protocols or allowing different and minimal estimation horizons in each local estimation problem will be subject of future research efforts.

7. Proofs

Since system (5) is time-invariant, for $N \geq \hat{r}^o = \max \{n^o_i\}$, Assumption 2 guarantees that

\begin{equation}
\|x_{t-N}^{[i]} - x_{t-N}^{[i]}\| \leq \gamma^o_i (\|w_k^{[i]} - w_k^{[i]}\|_{t-N+1}) + \gamma^o_i (\|y_k^{[i]} - y_k^{[i]}\|_{t-N+1})
\end{equation}

(34)

where $y_k^{[i]}$ and $y_k^{[i]}$ are the output sequences stemming from $(w_k^{[i]}, u_k^{[i]}, u_k^{[i]})$ and $(w_k^{[i]}, u_k^{[i]}, u_k^{[i]})$ with initial conditions $x_{t-N}^{[i]}$ and $x_{t-N}^{[i]}$, respectively.

**Lemma 1.** If Assumptions 3 and 4 hold then

\[ \sum_{k=t-N}^{t} L(\hat{w}_{k/z}, \hat{v}_{k/z}) \xrightarrow{t \to \infty} 0 \]  

(35)

**Proof.** The proof is similar to the one in [16]. For all $t \geq 0$, in view of (23) one has

\[ \Theta_j^{[i]} - \Theta_{j-1}^{[i]} \geq \sum_{k=t-N}^{t} L(\hat{w}_{k/z}, \hat{v}_{k/z}) \]  

(36)

By optimality $\Theta_j^{[i]} \leq \Xi\Xi\Xi_{t-N+1}^{[i]}((\{x_{t}(k, x_0)\})_{k=t-N+1}^{t}) \forall t > N$. The trajectory stemming from $x_{t}(t-N, x_0)$, $\hat{w}_{k}^{[i]} = \tilde{f}(x_{t}(k, x_0), x_{t}(k, x_0), u_k) - \tilde{f}(x_{t}(k, x_0), \hat{x}_{k-1}, u_k)$ for $k = t-N, \ldots, t-1$ and $\hat{w}_{k}^{[i]} = 0$ is $x_{t}(k, x_0)$ for $k = t-N+1, \ldots, t$, and hence it is feasible. Since $y_{t-N+1}$ corresponds to the deterministic system output (see Definition 1), it follows that $\hat{w}_{k}^{[i]} = y_{k} - h(x_{t}(k, x_0), u_k) = 0$ for all $k = t-N, \ldots, t$. Moreover, by optimality

\[ \Xi\Xi\Xi_{t-N+1}^{[i]}((\{x_{t}(k, x_0)\})_{k=t-N+1}^{t}) \leq J(t-N, t, x_{t}(t-N, x_0), \hat{w}_{k}, 0, \Gamma_{t-N}) \]

From (15), one has

\[ J(t-N, t, x_{t}(t-N, x_0), \hat{w}_{k}, 0, \Gamma_{t-N}) = \]

\[ = \sum_{k=t-N}^{t-1} L(\hat{w}_{k}, 0) + \Gamma_{t-N}(x_{t}(t-N, x_0); \hat{x}_{t-N-1}) \]

19
and in view of (24), $\Theta^*_t \leq \Xi_{[t-N_I+1]/t-1}(\{x_{\Sigma}(k, x_0)\}_{k=\gamma}^{t-1}$. We can iterate this procedure and prove that

$$\Theta^*_t \leq \Gamma_0(x_0; m_0) \leq \gamma_0(\|x_0 - m_0\|)$$

(37)

for all $t$, for any $x_0 \in \mathbb{X}$, where $m_0 \in \mathbb{X}$ is the prior estimate of $x_0$ and where $\gamma_0$ is a suitable $\mathcal{X}_0$ function, in view of (22a), which derives from Assumption 3.

Finally, from (36) the sequence $\Theta^*_t$ is increasing and from (37) it is bounded. Therefore $\Theta^*_t$ converges and, from (36), (35) follows.

**Proof of Theorem 1** Since Assumptions 3 and 4 hold, by Lemma 1 equation (35) is guaranteed. In view of (22a), it implies that

$$\max_{k \in \mathbb{N}} (\|\hat{\psi}_{k/t}\|, \|\hat{w}_{k/t}\|) \xrightarrow{t \to \infty} 0$$

(38)

Notice that, in the noiseless case ($w_{\Sigma}^d(k) = 0$ for all $k$ and $i = 1, \ldots, M$), for any $t$, the trajectory $x_{\Sigma}(t, x_0)$ is generated by the system

$$x_{\Sigma}(t + 1, x_0) = f(x_{\Sigma}(t + 1, x_0), x_{\Sigma}(t + 1, x_0), u_i)$$

(39)

and the output signal, for each sub-system, is

$$y^d_k = h^d(x^d_{\Sigma}(k, x_0), u^d_k)$$

and $\hat{y}_{k/t}^d = y_{k/t}^d - h^d(x_{k/t}^d, u_{k/t}^d)$. Recalling (34), we obtain that

$$\|x^d_{\Sigma}(t - N_I, x_0) - \|x^d_{\Sigma}(t - N, x_0)\| \leq \gamma^d_0 (\|w^d_{\Sigma}(k) - \hat{w}^d_{k/t}\|_{\mathcal{Y}_0} + \gamma^d_0 (\|x^d_{k/t}\|_{\mathcal{Y}_0} + \sum_{i \in \mathcal{M}} \gamma^d_0 (\|x^d_{k/t} - x^d_{\Sigma}(k, x_0)\|_{\mathcal{Y}_0} + \alpha^d_i)$$

(40)

From (38) we obtain that, for all $i = 1, \ldots, M$ there exists a positive sequence $\alpha^d_i$ satisfying $\alpha^d_i \xrightarrow{t \to \infty} 0$ such that

$$\|x^d_{\Sigma}(t - N_I, x_0) - x^d_{\Sigma}(t - N, x_0)\| \leq \sum_{j \in \mathcal{M}} \gamma^d_0 (\|x^d_{k/t} - x^d_{\Sigma}(k, x_0)\|_{\mathcal{Y}_0} + \alpha^d_i)$$

(41)

Recall that $u^d_{k/t} = \{x^d_{k/t}, j \in \mathcal{M}_I\}$ and, in the noiseless setting, $u^d_{k/t} = \{x^d_{\Sigma}(k, x_0), j \in \mathcal{M}_I\}$. For $k \geq t - N$, in view of (9b)

$$x^d_{k+1/t} = f^d(x^d_{k/t}, u^d_{k/t}, u^d_{k/t}) + w^d_{k/t}$$

(42)
Therefore one concludes that
\[ x_{\Sigma}^i(k+1, x_0) = f^i(x_{\Sigma}^i(k, x_0), u_k^i, x_k^i) \] (43)

Defining \( \Delta_{k_1/k_2}^i = \Delta_{k_1/k_2}^i - \Delta_{k_1}^i(x_0) \), we obtain, from (42) and (43)
\begin{align*}
\Delta_{k+1/t}^i &= f^i(x_{\Sigma}^i(k, x_0) + \Delta_{k/t}^i(x_{\Sigma}^i(k, x_0) + \Delta_{k/t-1}^i(j \in N, u_k^i)) \\
&- f^i(x_{\Sigma}^i(k, x_0), \{x_{\Sigma}^i(k, x_0)\})j \in N, u_k^i + \hat{w}_{k/t}^i
\end{align*}
(44)

From (44) and (6) (which derives from Assumption 1) it follows that:
\[ ||\Delta_{k+1/t}^i|| \leq l_{ii}||\Delta_{k/t}^i|| + \sum_{j \in N} l_{ij}||\Delta_{k/j-1}^i|| + ||\hat{w}_{k/t}^i|| \] (45)

Iterating equation (45), for \( k = t-N+1, \ldots, t \)
\[ ||\Delta_{k/t}^i|| \leq l_{ii}^{k-(t-N)}||\Delta_{t-N/t}^i|| + \sum_{j \in N} \sum_{r=0}^{k-(t-N)-1} l_{ij}^r||\Delta_{k-1-r/t-1}^i|| + \sum_{r=0}^{k-(t-N)-1} l_{ii}^r||\hat{w}_{k-1-r/t}^i|| \] (46)

Defining \( \delta_{k/t}^i = \max_{k \in [t-N+1: t]} ||\Delta_{k/t}^i|| \) and \( \alpha^w_{|t|} = \max_{k \in [t-N: t-1]} ||\hat{w}_{k/t}|| \) we can write, from (46)
\[ ||\Delta_{k/t}^i|| \leq l_{ii}^{k-(t-N)}||\Delta_{t-N/t}^i|| + \sum_{j \in N} l_{ij} \lambda(l_{ii}, k - (t-N)) \delta_{t-1}^j + \lambda(l_{ii}, k - (t-N)) \alpha^w_{|t|} \] (47)

where \( \lambda(\cdot, \cdot) \) is given in Definition 2. In view of (41), \( ||\Delta_{t-N/t}^i|| \leq \sum_{j \in N} \gamma_j \delta_{t-1}^j + \alpha^i_k \), and
\begin{align*}
||\Delta_{k/t}^i|| &\leq \sum_{j \in N} \left( l_{ii}^{k-(t-N)} \gamma_j \delta_{t-1}^j + l_{ij} \lambda(l_{ii}, k - (t-N)) \delta_{t-1}^j \right) \\
&+ \lambda(l_{ii}, k - (t-N)) \alpha^w_{|t|} + l_{ii}^{k-(t-N)} \alpha^i_k
\end{align*}
(48)

Therefore one concludes that
\begin{align*}
\delta_{k/t}^i &\leq \sum_{j \in N} \max_{k \in [t-N+1: t]} \left( l_{ii}^{k-(t-N)} \gamma_j \delta_{t-1}^j \right) \\
&+ l_{ij} \lambda(l_{ii}, k - (t-N)) \delta_{t-1}^j \] (49)
\begin{align*}
&+ \sum_{k \in [t-N+1: t]} \left( \lambda(l_{ii}, k - (t-N)) \alpha^w_{|t|} + l_{ii}^{k-(t-N)} \alpha^i_k \right)
\end{align*}
We define
\[ \tilde{\alpha}_i[t] = \max_{k \in \{t-N+1, \ldots, t\}} \left( \lambda(l_{ii}, k - (t-N)) \alpha^{w_i}_i + t_{ii}^{k-(t-N)} \alpha^{t_i}_i \right) \]
which, in view of (38), is an asymptotically vanishing term. Furthermore, in view of Definition 3, we can write (49) as
\[ \delta_i[t] \leq \sum_{j \in \mathcal{N}_i} \tilde{\gamma}_{ij} \delta_j[t-1] + \tilde{\alpha}_i[t] \]
(50)
The stability of the system of interconnected equations (50) can be analyzed by means of the small gain condition given in [2].

Define a vector \( \delta \in \mathbb{R}^M \), with components \( \delta_i \geq 0 \), \( i = 1, \ldots, M \). Since, by definition, for all \( i \), the \( i \)-th component of \( \tilde{\Gamma}(\delta) \) does not depend on \( \delta_i \) and is a \( \mathcal{K}_\infty \) function of \( \delta_i \), \( j \in \mathcal{N}_i \), it is easy to see that the system of equations (50) is asymptotically stable (i.e., \( \delta_i[t] \to 0 \) as \( t \to \infty \) for all \( i = 1, \ldots, M \)) if the map \( \tilde{\Gamma} \) satisfies the small gain condition (30).

**Proof of Corollary 1**

If the system partition induces a DAG and since Assumption 1 holds, \( \mathcal{L} \), in (6), is lower triangular, after a suitable permutation of the subsystem indexes. Therefore, without loss of generality, we have \( l_{ij} = 0 \) if \( j > i \). Also, since Assumption 2 holds, similar arguments apply to \( \gamma_j \) in (8). According to (28), this gives rise to a map \( \tilde{\Gamma}(\delta) \) whose \( i \)-th element depends only upon \( \delta_j \), with \( j < i \), for all \( i = 1, \ldots, M \).

In view of the structure of \( \tilde{\Gamma} \), the small gain condition is verified. Note that any admissible \( \delta \) satisfies the following: there exists an index \( i \) such \( \delta_i = 0 \) for all \( j < i \) and \( \delta_i > 0 \) (if \( i = 1 \) this corresponds only to the condition \( \delta_1 > 0 \)). In view of its structure, the \( i \)-th entry of \( \tilde{\Gamma} \) is equal to zero i.e., \( \tilde{\Gamma}_i = 0 \). Therefore, for all \( \delta \), there exists an index \( i \) such that \( \tilde{\Gamma}(D(\delta))_i < \delta_i \). This corresponds to condition (30). Being Assumptions 1, 2, 3, 4 verified, we resort to Theorem 1 which guarantees asymptotic convergence of NPMHE.

**References**


