Partition-based Pareto-Optimal State Prediction Method for Interconnected Systems using Sensor Networks

Yilun Zhou, Francesca Boem and Thomas Parisini

***Abstract*— In this paper a novel partition-based state prediction method is proposed for interconnected stochastic systems using sensor networks. Each sensor locally computes a prediction of the state of the monitored subsystem based on the knowledge of the local model and the communication with neighboring nodes of the sensor network. The prediction is performed in a distributed way, not requiring a centralized coordination or the knowledge of the global model. Weights and parameters of the state prediction are locally optimized in order to minimise at each time-step bias and variance of the prediction error by means of a multi-objective Pareto optimization framework. Individual correlations between the state, the measurements, and the noise components are considered, thus assuming to have in general unequal weights and parameters for each different state component. No probability distribution knowledge is required for the noise variables. Simulation results show the effectiveness of the proposed method.**

I. Introduction

One fundamental application of sensor networks is the estimation and prediction of the state of Large-Scale Systems (LSSs). This problem finds application for several activities, e.g., target tracking, environmental monitoring, industrial plants process control. Nevertheless, there is an increasing demand for innovative methods for the monitoring of interconnected LSSs. This problem is challenging and centralized solutions are usually not feasible due to communication and computation constraints. The technological availability of cheap sensors rises new challenges on how to use the collected information. In this paper, we address the problem of partition-based state prediction using sensor networks to monitor interconnected systems in a distributed manner.

Partition-based state estimation is an active research area, where each local agent estimates part of the global system's state vector using only local model information and communicating only with neighboring agents (see [1], [2], [3], [4], [5], [6]). In [1] and [2], Kalman-consensus based distributed estimators are proposed, while [6] introduces a Kalman-filter based distributed estimator without consensus strategy for non-overlapping interconnected subsystems. None of these works use sensor networks for the estimation task allowing more than one sensor to monitor the same state vector. To the authors' knowledge, this is the first contribution proposing a partition-based prediction method using sensor networks. On the other hand, many works propose Kalman-based and Kalman-consensus filtering approaches using sensor networks (see [7], [8], [9], [10], [11], [12]) to estimate the entire state vector exploiting the knowledge of the global model. The advantages of a partition-based approach are manifold: each sensor needs only local model information, potentially dealing with more scalable architectures; the computation cost at each node is reduced; as in distributed approaches, only communication with neighboring nodes is required.

In this paper, a multi-objective optimization problem is locally solved to jointly minimize mean and variance of the prediction error. With respect to our previous works ([13], [14], [15]), it is the first time that we consider the partition-based estimation problem using sensor networks. In [13], a noisy signal is estimated using a sensor network, while [15] proposes a distributed state prediction method, where each sensor estimates the entire state vector based on the model of the global system. Each sensor may communicate both with other sensors monitoring the same subsystem, and with sensors measuring the state of neighboring subsystems. There are no assumptions on the communication network topology, apart from connectedness. The time-varying weights to consider the available information at each time step are designed in a Pareto-optimal architecture. We consider correlations between the local state and neighboring systems'states, thus dealing with a more challenging scenario with respect to our previous works ([13], [14], [15]). Convergence conditions of the estimation error are provided. The on-line computation of the time-varying weights allows to consider also the transient performance, together with the asymptotic performance, differently from other methods [2], [6], [5], where only asymptotic performance is investigated.

Notation. Given a stochastic variable x, $\mathbb{E}x$ represents its expected value. By $\mathbf{1}_s$, $\mathbf{0}_s$ and I_s we denote the vectors $(1,\dots,1)^\top$, $(0,\dots,0)^\top$ and the identity matrix with appropriate size s, respectively. Given a vector v, we denote $\mathrm{diag}(v)$ the diagonal matrix whose diagonal entries are the elements of v. $|\cdot|$ denotes the cardinality of a set and $\|\cdot\|$ the spectral norm of a matrix. Finally, $\otimes$ denotes the Kronecker product, $\bigoplus$ the direct sum of matrices and the operator $\circ$ represents the component-by-component product.

II. Problem formulation

The monitored system is composed of (or can be decomposed in) N interconnected subsystems, each modeled as

$$\Sigma_I \ : \ x_I(t+1) = A_{II}x_I(t) + \sum_{J\in\mathcal{P}_I} A_{IJ}x_J(t) + w_I(t), \quad (1)$$

F. Boem and Y. Zhou are with the Dept. of Electrical and Electronic Engineering at the Imperial College London, UK. (f.boem@imperial.ac.uk, yilun.zhou11@imperial.ac.uk)

T. Parisini is with the Dept. of Electrical and Electronic Engineering at the Imperial College London, UK, and with the Dept. of Engineering and Architecture at the University of Trieste, Italy (t.parisini@gmail.com)

This work has been partially supported by the research project "Stability and Control of Power Networks with Energy Storage (STABLE-NET)," funded by the RCUK Energy Programme (contract no: EP/L014343/1).

with $I = 1, \dots, N$, where $x_I \in \mathbb{R}^{m_I}$ denotes the local state vector, $w_I \in \mathbb{R}^{m_I}$ represents process disturbances, $\mathcal{P}_I$ is a set collecting parents of subsystem Σ_I, that is, the subsystems Σ_J whose state x_J influence the dynamics of Σ_I. Matrix A_{II} describes local dynamics, while A_{IJ} models the dynamic coupling between Σ_I and Σ_J, $J \in \mathcal{P}_I$.

The system is monitored by a sensor network $\mathcal{S}$, composed of n sensors. Each subsystem Σ_I, $I = 1, \dots, N$, is monitored by the set of sensors $\mathcal{S}_I$, composed of n_I sensors. Each sensor $i \in \mathcal{S}_I$ measures the state x_I, according to the following measurement equation:

$$y_I^i(t) = x_I(t) + v_I^i(t), \tag{2}$$

where $y_I^i \in \mathbb{R}^{m_I}$, denotes the measurements vector taken by sensor $i \in \mathcal{S}_I$ and $v_I^i \in \mathbb{R}^{m_I}$ is the measurement noise.

Assumption 1: We assume to know the mean $\bar{w}_I$ and the covariance matrix $\Sigma_{wI}(t)$ of the process noise vector $w_I(t)$; furthermore, $v_I^i \in \mathbb{R}^{m_I}$ is a zero-mean noise vector, with $\Sigma_{v_I^i}$ its covariance matrix.

The sensors exchange information by means of a communication network, modeled as an undirected graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$, where $\mathcal{S}$ is the set of the nodes (the sensors) and $\mathcal{E}$ is the set of the edges connecting the nodes.

According to the graph $\mathcal{G}$, each sensor $i \in \mathcal{S}_I$ may communicate with two different subsets of nodes (see also the example in Fig. 1, Section V):

- some neighboring sensors in $\mathcal{S}_I$, which we denote $\mathcal{N}_I^i := \{l \in \mathcal{S}_I : (l,i) \in \mathcal{E}\} \cup \{i\}$, including the set of neighbors of node $i \in \mathcal{S}$ monitoring the same subsystem Σ_I, plus the node i itself.
- some neighbouring sensors in $\mathcal{S}_J$, $J \in \mathcal{P}_I$, collected in the sets $\mathcal{N}_{IJ}^i := \{l \in \mathcal{S}_J, J \in \mathcal{P}_I : (l,i) \in \mathcal{E}\}$, monitoring parents of subsystem Σ_I.

III. Distributed state prediction

In this paper each node i of the sensor network implements a two steps dynamic estimator: a filtering step and a prediction step[1]. First, by communicating with neighboring nodes in $\mathcal{N}_I^i$, it filters the measurement noise in a consensus-like fashion by computing a linear combination of its own and neighbors' available measurements and predictions:

$$\bar{x}_I^i(t) = \sum_{l \in \mathcal{N}_I^i} [k_I^{i,l}(t)\hat{x}_I^l(t) + h_I^{i,l}(t) y_I^l(t)], \tag{3}$$

where $k_I^{i,l}(t)$ and $h_I^{i,l}(t) \in \mathbb{R}^{m_I \times m_I}$ are diagonal matrices collecting the time-varying filter weights. The objective of this first step, similarly as in [15], is for each node to reduce its own measurement uncertainty, without the use of centralized coordination.

After the consensus-filtering step, each node implements a model-based prediction. In this second phase each node $i \in \mathcal{S}_I$ communicates with neighboring nodes $\mathcal{N}_{IJ}^i$, $J \in \mathcal{P}_I$, to consider also the coupling influence. The one-step-ahead prediction is computed as

$$\begin{aligned}\hat{x}_I^i(t+1) = A_{II}\bar{x}_I^i(t) + \bar{w}_I(t) + \lambda_I'^{\,i}(t)(\hat{x}_I^i(t) - \bar{x}_I^i(t)) \\ + \sum_{J \in \mathcal{P}_I} A_{IJ} \sum_{j \in \mathcal{N}_{IJ}^i} \omega_{IJ}^{i,j}(t)\hat{x}_J^j(t),\end{aligned} \tag{4}$$

where $\lambda_I'^{\,i}(t) = \lambda_I^i(t)A_{II}$, being $\lambda_I^i(t) \in \mathbb{R}^{m_I \times m_I}$ and $\omega_{IJ}^{i,j}(t) \in \mathbb{R}^{m_J \times m_J}$ diagonal matrices collecting the time-varying filter parameters.

The goal is to design $k_I^{i,l}(t)$, $h_I^{i,l}(t)$, $\lambda_I^i(t)$ and $\omega_{IJ}^{i,j}(t)$, for each $I = 1, \dots, N$, $i = 1, \dots, n_I$, $l = 1, \dots, m_I$, $J \in \mathcal{P}_I$, so to minimize at each time step bias and variance of the global prediction errors.

A. Local estimation and prediction errors

Let us define the local filtering error $\bar{e}_I^i(t) = \bar{x}_I^i(t) - x_I^i(t)$ and the local prediction error

$$\hat{E}_I^i(t) = \hat{x}_I^i(t) - x_I^i(t).$$

We rewrite Eqs. (3) and (4) as:

$$\begin{aligned}\bar{x}_I^i(t) &= \kappa_I^i(t)\hat{x}_I^{i\mathrm{reg}}(t) + \eta_I^i(t) y_I^{i\mathrm{reg}}(t) \\ \hat{x}_I^i(t+1) &= A_{II}\bar{x}_I^i(t) + \lambda_I^i(t)(A_{II}\hat{x}_I^i(t) - A_{II}\bar{x}_I^i(t)) \\ &\quad + \bar{w}_I(t) + A_I^{i\mathrm{nei}}\omega_I^i(t)\hat{x}_I^{i\mathrm{nei}}(t),\end{aligned} \tag{5}$$

where $\hat{x}_I^{i\mathrm{reg}}$ and $y_I^{i\mathrm{reg}}$ are two column vectors collecting the prediction vectors and the measurements vectors (respectively) available at node i related to subsystem Σ_I, ordered according to their indexes $i_1 < \dots < i_{N_I^i}$:

$$\hat{x}_I^{i\mathrm{reg}} = (\hat{x}_I^{i_1\top}, \dots, \hat{x}_I^{i_{N_I^i}\top})^\top, \quad y_I^{i\mathrm{reg}} = (y_I^{i_1\top}, \dots, y_I^{i_{N_I^i}\top})^\top$$

with N_I^i being the cardinality of the set $\mathcal{N}_I^i$. Moreover, $\hat{x}_I^{i\mathrm{nei}}$ is a column vector collecting the prediction vectors available at node i related to subsystems Σ_J, $J \in \mathcal{P}_I$, ordered according to their indexes. Furthermore, $\kappa_I^i(t)$ and $\eta_I^i(t)$ are the time varying row block matrices $\in \mathbb{R}^{m_I \times N_I^i m_I}$ collecting matrices $k_I^{i,l}$ and $h_I^{i,l}$ respectively; $A_I^{i\mathrm{nei}}$ is a row block matrix collecting matrices A_{IJ}, $J \in \mathcal{P}_I$. Finally, $\omega_I^i(t)$ is a block matrix collecting on the diagonal blocks matrices $\omega_{IJ}^i(t)$, which are row block matrices collecting $\omega_{IJ}^{i,j}(t)$ according to the order followed in $\hat{x}_I^{i\mathrm{nei}}$.

To derive the optimization problem in Section IV, the following constraints are introduced: at each time t

$$\begin{aligned}(\kappa_I^i(t) + \eta_I^i(t))\mathbf{1}_{N_I^i m_I} &= \mathbf{1}_{m_I}, \\ \omega_I^i(t)\mathbf{1}_{p_I^i} &= \mathbf{1}_{m_I^i},\end{aligned} \tag{6}$$

being N_{IJ}^i the cardinality of set $\mathcal{N}_{IJ}^i$ and $p_I^i = \sum_{J \in \mathcal{P}_I} N_{IJ}^i m_J$, $m_I^i = \sum_{J \in \mathcal{P}_I} m_J$.

These are needed so that the following expressions hold:

$$\kappa_I^i(t)x_I^E(t) + \eta_I^i(t)x_I^E(t) = x_I^E(t)$$

being x_I^E a column vector repeating N_I^i times the state vector x_I, and

$$\omega_{IJ}^i(t)x_J^{iE}(t) = x_J^{iE}(t)$$

[1]It is worth noting that in the literature it is quite a common choice to present prediction methods by means of a 2-steps strategy made of a merging-update step and a prediction step (see, for instance [8], [9], [16]).

being x_J^{iE} a column vector repeating N_{IJ}^i times the state vector x_J, for each $J \in \mathcal{P}_I$. We use these expressions to derive the local filtering error and the local prediction error:

$$\bar{e}_I^i(t) = \kappa_I^i(t)\hat{\epsilon}_I^i(t) + \eta_I^i(t)v_{\epsilon_I^i}(t), \tag{7}$$

$$\begin{aligned}\hat{E}_I^i(t+1) =& A_{II}(I-\lambda_I^i(t))\kappa_I^i(t)\hat{\epsilon}_I^i(t) + A_{II}\lambda_I^i(t)\hat{E}_I^i(t)\\ &+ A_{II}(I-\lambda_I^i(t))\eta_I^i(t)v_{\epsilon_I^i}(t) - w_I(t)\\ &+ \bar{w}_I + A_I^{i\text{nei}}\omega_I^i(t)\hat{\epsilon}_I^{i\text{nei}}(t),\end{aligned} \tag{8}$$

where $\hat{\epsilon}_I^i$ collects the prediction error vectors available at node i and $v_{\epsilon_I^i}$ collects the measurement noise vectors related to the measurements available at node i, both related to subsystem Σ_I and ordered following their indexes; $\hat{\epsilon}_I^{i\text{nei}}$ collects the prediction error vectors available at node i related to subsystems Σ_J, $J \in \mathcal{P}_I$.

We derive the expressions of the bias and the variance for the local filtering and prediction errors, given in (7) and (8), respectively. The expected values can be computed as

$$\begin{aligned}\mathbb{E}\bar{e}_I^i(t) &= \kappa_I^i(t)\mathbb{E}\hat{\epsilon}_I^i(t),\\ \mathbb{E}\hat{E}_I^i(t+1) &= A_{II}(I-\lambda_I^i(t))\kappa_I^i(t)\mathbb{E}\hat{\epsilon}_I^i(t)\\ &\quad + A_{II}\lambda_I^i(t)\mathbb{E}\hat{E}_I^i(t) + A_I^{i\text{nei}}\omega_I^i(t)\mathbb{E}\hat{\epsilon}_I^{i\text{nei}}(t)\,,\end{aligned} \tag{9}$$

The variance can be computed as

$$\begin{aligned}&\mathbb{E}[(\hat{E}_I^i(t+1) - \mathbb{E}\hat{E}_I^i(t+1))(\hat{E}_I^i(t+1) - \mathbb{E}\hat{E}_I^i(t+1))^\top]\\ =&W_{1i}(t)\Gamma_{\hat{\epsilon}_I^i}(t)W_{1i}(t)^\top + W_{2i}(t)\Sigma_{v_{\hat{\epsilon}_I^i}}W_{2i}(t)^\top\\ &+ W_{3i}(t)\Gamma_{\hat{\epsilon}_I^{i\text{nei}}}(t)W_{3i}(t)^\top + \Sigma_{w_I}(t)\end{aligned} \tag{10}$$

being $\text{Cov}(\hat{\epsilon}_I^i(t), v_{\hat{\epsilon}_I^i}(t)) = 0$, $\text{Cov}(\hat{\epsilon}_I^i(t), w_I(t)) = 0$, $\text{Cov}(v_{\hat{\epsilon}_I^i}(t), w_I(t)) = 0$ and $\text{Cov}(\hat{\epsilon}_I^i(t), \hat{\epsilon}_I^{i\text{nei}}(t)) = 0$, where

$$W_{1i}(t) = A_{II}[(I-\lambda_I^i(t))\kappa_I^i(t) + \lambda_I^i(t)Z_I^i], \tag{11}$$

$$W_{2i}(t) = A_{II}(I-\lambda_I^i(t))\eta_I^i(t), \tag{12}$$

$$W_{3i}(t) = A_I^{i\text{nei}}\omega_I^i(t), \tag{13}$$

$$\Gamma_{\hat{\epsilon}_I^i}(t) = \mathbb{E}[(\hat{\epsilon}_I^i(t) - \mathbb{E}\hat{\epsilon}_I^i(t))(\hat{\epsilon}_I^i(t) - \mathbb{E}\hat{\epsilon}_I^i(t))^\top], \tag{14}$$

$$\Gamma_{\hat{\epsilon}_I^{i\text{nei}}}(t) = \mathbb{E}[(\hat{\epsilon}_I^{i\text{nei}}(t) - \mathbb{E}\hat{\epsilon}_I^{i\text{nei}}(t))(\hat{\epsilon}_I^{i\text{nei}}(t) - \mathbb{E}\hat{\epsilon}_I^{i\text{nei}}(t))^\top] \tag{15}$$

Z_I^i is a $m_I \times m_I N_I^i$ block vector, where the block located at the position corresponding to the i-th index in the set $\mathcal{N}_I^i$, is the identity matrix I_{m_I}; all the other blocks are equal to 0. $\Sigma_{v_{\hat{\epsilon}_I^i}}$ is the measurement noise covariance matrix, including correlations between neighboring sensors.

B. Estimation error stability

Some local conditions on the time-varying weights are introduced to guarantee that the expected value of the global prediction error can converge to zero. For the sake of simplicity, we omit the dependence on t of the matrices.

Assumption 2: We assume that matrix A, describing the dynamics of the global system, satisfies $\|A\|_\infty < 1$.

Proposition 3.1: Under Assumption 2, the following local conditions are sufficient to guarantee the asymptotic stability of the expected value of the local prediction error (9) at each node i. For each r-th row of κ_I^i and of λ_I^i, and each r'-th row of $\omega_{I,J}^i$, with $I = 1,\dots,N$, $i = 1,\dots,n_I$, $r = 1,\dots,m_I$ and $r' = 1,\dots,m_J$:

$$\begin{aligned}&\sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right| < \frac{1}{\|A_I^{Eg}\|_\infty}\\ &\frac{-\frac{1}{\|A_I^{Eg}\|_\infty} + \sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|}{1+\sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|} < \lambda_I^{i^r} < \frac{\frac{1}{\|A_I^{Eg}\|_\infty} + \sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|}{1+\sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|}\\ &\sum_{l=1}^{n_J}\left|\omega_{I,J}^{i,l^{r'}}\right| < \frac{1}{\|A_I^{Eg}\|_\infty}\,, J \in \mathcal{P}_I\,,\end{aligned} \tag{16}$$

where A_I^{Eg} collects the rows of matrix A related to Σ_I. The proof is omitted due to length constraints.

Remark. Assumption 2 is used in the proof to simplify some inequality relations. Similar convergence sufficient conditions can be derived for different matrix A cases. We omit the analysis due to space constraints.

IV. The optimization problem

The goal of the proposed distributed method is to predict the local state minimizing the bias and variance of the prediction error at each sensor at each time step. To do that, we propose that each sensor at each step computes the optimal time-varying weights by solving a multi objective optimization problem, where the first objective is the squared bias and the second objective is the variance of the prediction error. We define the following Pareto optimization problem:

$$\min_{\kappa_I^i(t),\eta_I^i(t),\lambda_I^i(t),\omega_I^i(t)} \text{tr}\,[\rho_I^i(B_I^i)^2 + (1-\rho_I^i)V_I^i] \tag{17a}$$

$$\text{s.t. } (\kappa_I^i(t) + \eta_I^i(t))\mathbf{1}_{N_I^i m_I} = \mathbf{1}_{m_I} \tag{17b}$$

$$\omega_I^i(t)\mathbf{1}_{p_I^i} = \mathbf{1}_{m_I^i} \tag{17c}$$

$$\text{convergence conditions Eq.(16)} \tag{17d}$$

where $0 \le \rho_I^i \le 1$ is the Pareto parameter, $B_I^i = \mathbb{E}\hat{E}_I^i(t+1)$ is the prediction error bias given in Eq. (9), $V_I^i = \mathbb{E}[(\hat{E}_I^i(t+1) - \mathbb{E}\hat{E}_I^i(t+1))(\hat{E}_I^i(t+1) - \mathbb{E}\hat{E}_I^i(t+1))^\top]$ is the variance of the prediction error given in Eq. (10).

A. The approximated problem

We briefly analyze the convergence conditions (16). Since the absolute value would make the problem more difficult to solve, we use the more restrictive conditions

$$\begin{cases}\kappa_I^{i\top}(t)\mathbf{1}_{m_I} \ge \mathbf{0}_{N_I^i m_I}\\ \kappa_I^i(t)\mathbf{1}_{N_I^i m_I} < \dfrac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I},\end{cases} \tag{18}$$

$$\begin{cases}\omega_I^{i\top}(t)\mathbf{1}_{m_I^i} \ge \mathbf{0}_{p_I^i}\\ \omega_I^i(t)\mathbf{1}_{p_I^i} < \dfrac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I^i},\end{cases} \tag{19}$$

$$0 \le \lambda_I^i(t) \le I_{m_I}, \tag{20}$$

implying conditions (16), by noting that

$$\frac{-\frac{1}{\|A_I^{Eg}\|_\infty}+\sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|}{1+\sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|}<0,$$

$$\frac{\frac{1}{\|A_I^{Eg}\|_\infty}+\sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|}{1+\sum_{l=1}^{n_I}\left|k_I^{i,l^r}\right|}>1.$$

Based on (9) and (10), we rewrite the Pareto optimization problem (17) by introducing the following terms which are data of the problem or can be computed on-line empirically:

$$M_I^i(\rho_I^i)=(1-\rho_I^i)\Gamma_{\hat{\epsilon}_I^i},$$

$$M_I^{i\text{nei}}(\rho_I^i)=(1-\rho_I^i)\Gamma_{\hat{\epsilon}_I^{i\text{nei}}},$$

$$R_I^i(\rho_I^i)=\rho_I^i(\Upsilon_{I1}^i\mathbb{E}\hat{\epsilon}_I^i+\Upsilon_{I2}^i\mathbb{E}\hat{\epsilon}_I^{i\text{nei}})(\Upsilon_{I1}^i\mathbb{E}\hat{\epsilon}_I^i+\Upsilon_{I2}^i\mathbb{E}\hat{\epsilon}_I^{i\text{nei}})^\top,$$

$$S_I^i(\rho_I^i)=(1-\rho_I^i)\Sigma_{v_{\hat{\epsilon}_I^i}},$$

with Υ_{I1}^i and Υ_{I2}^i are both column block matrices, denoted by $\Upsilon_{I1}^i=\text{col}(I_{m_I N_I^i},\mathbf{0}_{p_I^i}\otimes\mathbf{1}_{m_I N_I^i}^\top)$ and $\Upsilon_{I2}^i=\text{col}(\mathbf{0}_{p_I^i}^\top\otimes\mathbf{1}_{m_I N_I^i},I_{p_I^i})$, respectively. The dependence on the time is removed for simplicity. We obtain

$$\begin{aligned}\min_{\kappa_I^i,\eta_I^i,\lambda_I^i,\omega_I^i}\ &\text{tr}[A_I^{i\text{nei}}\omega_I^i M_I^{i\text{nei}}(\rho_I^i)\omega_I^{i\top}A_I^{i\text{nei}\top}+(1-\rho_i)\Sigma_w\\&+A_{II}[(I-\lambda_I^i)\kappa_I^i+\lambda_I^i Z_I^i]M_I^i(\rho_I^i)[(I-\lambda_I^i)\kappa_I^i+\lambda_I^i Z_I^i]^\top A_{II}^\top\\&+(A_{II}[(I-\lambda_I^i)\kappa_I^i+\lambda_I^i Z_I^i]\Upsilon_{I1}^{i\top}+A_I^{i\text{nei}}\omega_I^i\Upsilon_{I2}^{i\top})R_I^i(\rho_I^i)\\&\times(A_{II}[(I-\lambda_I^i)\kappa_I^i+\lambda_I^i Z_I^i]\Upsilon_{I1}^{i\top}+A_I^{i\text{nei}}\omega_I^i\Upsilon_{I2}^{i\top})^\top\\&+A_{II}(I-\lambda_I^i)\eta_I^i S_I^i(\rho_I^i)\eta_I^{i\top}(I-\lambda_I^i)^\top A_{II}^\top]\end{aligned}\tag{21a}$$

$$\text{s. t. }(\kappa_I^i+\eta_I^i)\mathbf{1}_{N_I^i m_I}=\mathbf{1}_{m_I}\tag{21b}$$

$$\omega_I^i\mathbf{1}_{p_I^i}=\mathbf{1}_{m_I^i}\tag{21c}$$

$$\kappa_I^{i\top}\mathbf{1}_{m_I}\geq\mathbf{0}_{N_I^i m_I}\tag{21d}$$

$$\kappa_I^i\mathbf{1}_{N_I^i m_I}<\frac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I}\tag{21e}$$

$$\omega_I^{i\top}\mathbf{1}_{m_I^i}\geq\mathbf{0}_{p_I^i}\tag{21f}$$

$$\omega_I^i\mathbf{1}_{p_I^i}<\frac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I^i}\tag{21g}$$

$$\lambda_I^i\mathbf{1}_{m_I}\leq\mathbf{1}_{m_I}\tag{21h}$$

$$\lambda_I^i\mathbf{1}_{m_I}\geq\mathbf{0}_{m_I}\tag{21i}$$

Note that problems (17) and (21) have the same objective function, but (21) is constrained by more restrictive conditions. This gives an optimization problem which is convex since the objective function has a quadratic form and M_I^i, $M_I^{i\text{nei}}$, R_I^i and S_I^i are positive definite matrices. Coherently, we can use Lagrangian duality to solve the problem.

B. The optimal weights

By means of the Karush Kuhn Tucker (KKT) conditions (see Appendix), which are both sufficient and necessary conditions for optimality, it is possible to derive the optimal values for the decisional variables $\kappa_I^i(t)$, $\eta_I^i(t)$, $\lambda_I^i(t)$ and $\omega_I^i(t)$ of Problem (21). Define the $m_I N_I^i\times 1$ vector $\kappa_I^{i\text{vec}}(t)=\kappa_I^{i\top}(t)\mathbf{1}_{m_I}$, collecting all the diagonals of each block in $\kappa_I^i(t)$ on a column vector. Similarly, we denote $\eta_I^{i\text{vec}}(t)=\eta_I^{i\top}(t)\mathbf{1}_{m_I}$, the $p_I^i\times 1$ vector $\omega_I^{i\text{vec}}(t)=\omega_I^{i\top}(t)\mathbf{1}_{m_I^i}$ and the $m_I\times 1$ vector $\lambda_I^{i\text{vec}}(t)=\lambda_I^{i\top}(t)\mathbf{1}_{m_I}$. Let introduce the dual variables $\xi_{I1}^i(t)$, $\xi_{I2}^i(t)$, $\xi_{I4}^i(t)$ and $\nu_{I1}^i(t)$, which are $m_I\times 1$ vectors, the $m_I N_I^i\times 1$ vector $\xi_{I3}^i(t)$, the $p_I^i\times 1$ vector $\xi_{I5}^i(t)$, and the $m_I^i\times 1$ vectors $\xi_{I6}^i(t)$ and $\nu_{I2}^i(t)$. For the sake of notation simplicity, in the following we omit the dependence on t and on ρ_I^i.

Proposition 4.1: The solution for problem (21), for each node i in subsystem I, is given by $\xi_{I1}^i=0$, $\xi_{I2}^i=0$, $\xi_{I3}^i=0$, $\xi_{I4}^i=0$, $\xi_{I5}^i=0$ and $\xi_{I6}^i=0$,

$$\kappa_I^{i\text{vec}}=-P_7^{-1}(P_2P_4^{-1}I_\omega P_5^{-1}\mathbf{1}_{m_I^i}+I_\kappa\nu_{I1}^i+P_2P_6Z_\lambda^2+Z_\lambda^1),\tag{22a}$$

$$\eta_I^{i\text{vec}}=-P_3^{-1}I_\kappa\nu_{I1}^i,\tag{22b}$$

$$\omega_I^{i\text{vec}}=P_6P_2^\top\kappa_I^{i\text{vec}}+P_6Z_\lambda^2+P_4^{-1}I_\omega P_5^{-1}\mathbf{1}_{m_I^i},\tag{22c}$$

$$\begin{aligned}\lambda_I^{i\text{vec}}=&((A_{II}^\top A_{II})^\top\circ(\eta_I^iS_I^i\eta_I^{i\top}+(Z_I^i-\kappa_I^i)F_I^i(Z_I^i-\kappa_I^i)^\top))^{-1}\\&\times\text{diag}^{-1}([(\kappa_I^i-Z_I^i)F_I^i\kappa_I^{i\top}+\eta_I^iS_I^i\eta_I^{i\top}]A_{II}^\top A_{II}\\&+(\kappa_I^i-Z_I^i)Q_I^i\omega_I^{i\top}A_I^{i\text{nei}\top}A_{II}),\end{aligned}\tag{22d}$$

with

$$\begin{aligned}Z_\lambda^1=&([N_I^iZ_I^{i\top}\lambda_I^{i\top}A_{II}^\top A_{II}(I-\lambda_I^i)]\circ I_\kappa)\mathbf{1}_{m_I},\\Z_\lambda^2=&([Q_I^{i\top}Z_I^{i\top}\lambda_I^{i\top}A_{II}^\top A_I^{i\text{nei}}]\circ I_\omega)\mathbf{1}_{m_I^i},\\\nu_{I1}^i=&-P_8^{-1}(\mathbf{1}_{m_I}+I_\kappa^\top P_7^{-1}[P_2P_4^{-1}I_\omega P_5^{-1}\mathbf{1}_{m_I^i}+P_2P_6Z_\lambda^2\\&+Z_\lambda^1]),\ P_1=2\mathbf{D}_I^\top\circ F_I^i,P_2=2\mathbf{D}_{IJ}^\top\circ Q_I^i,\\P_3=&2\mathbf{D}_I^\top\circ S_I^i,P_4=2\mathbf{D}_J^\top\circ F_I^{i\text{nei}},P_5=I_\omega^\top P_4^{-1}I_\omega,\\P_6=&P_4^{-1}I_\omega P_5^{-1}I_\omega^\top P_4^{-1}-P_4^{-1},P_7=P_1+P_2P_6P_2^\top,\\P_8=&I_\kappa^\top P_7^{-1}I_\kappa+I_\kappa^\top P_3^{-1}I_\kappa,\end{aligned}$$

where $I_\kappa=\mathbf{1}_{N_I^i}\otimes I_{m_I}$, $I_\lambda=I_{m_I}$, $I_\omega=\bigoplus_{J\in\mathcal{P}_I}(\mathbf{1}_{N_{IJ}^i}\otimes I_{m_J})$, $D_I=(I-\lambda_I^i)^\top A_{II}^\top A_{II}(I-\lambda_I^i)$, $D_{IJ}=A_I^{i\text{nei}\top}A_{II}(I-\lambda_I^i)$, $D_J=A_I^{i\text{nei}\top}A_I^{i\text{nei}}$, $\mathbf{D}_I=\mathbf{1}_{N_I^i}\mathbf{1}_{N_I^i}^\top\otimes D_I$, $\mathbf{D}_{IJ}=\mathbf{1}_{N_I^i}^\top\otimes(I_\omega D_{IJ})$, $\mathbf{D}_J=I_\omega D_J I_\omega^\top$, $F_I^i(\rho_I^i)=M_I^i(\rho_I^i)+\Upsilon_{I1}^{i\top}R_I^i(\rho_I^i)\Upsilon_{I1}^i$, $F_I^{i\text{nei}}(\rho_I^i)=M_I^{i\text{nei}}(\rho_I^i)+\Upsilon_{I2}^{i\top}R_I^i(\rho_I^i)\Upsilon_{I2}^i$, and $Q_I^i(\rho_I^i)=\Upsilon_{I1}^{i\top}R_I^i(\rho_I^i)\Upsilon_{I2}^i$.

The proof is omitted due to space constraints.

In order to use the result in Prop. 4.1, each node implements Algorithm 1 to find optimal values κ_I^{i*}, η_I^{i*}, ω_I^{i*} and λ_I^{i*}. At each step, we verify that the obtained values satisfy conditions (16). Define a small positive constant ϵ_I^i, the computation is stopped until difference between current value of λ_I^i and the updated one is smaller than ϵ_I^i.

Remark. Algorithm 1 has complexity $O(\text{N}_{\text{Iter1}}(2(m_I\times|\mathcal{N}_I^i|)^3+2(m_I)^3+(p_I^i)^3+(m_I^i)^3))$, given by the computation of matrices inverse, where N_{Iter1} is the number of iterations. In the simulation, N_{Iter1} at each time step is lower than 10. Furthermore, the computation of a covariance matrix is required to compute the estimates of $\Gamma_{\hat{\epsilon}_I^i}$, $\Gamma_{\hat{\epsilon}_I^{i\text{nei}}}$, $\mathbb{E}\hat{\epsilon}_I^i$, and $\mathbb{E}\hat{\epsilon}_I^{i\text{nei}}$: the complexity is $O(\text{Table}_{\text{size}}\log(\text{Table}_{\text{size}}))$, where the $\text{Table}_{\text{size}}$ is the size of a look-up table used to speed up

the computation of a quadratically constrained least-square problem [17]. We set $\text{Table}_{\text{size}} = 100$.

Algorithm 1 Optimal weights computation

Set $\lambda_I^{i+} = I_{m_I}/2$, ϵ_I^i

repeat

$\lambda_I^i = \lambda_I^{i+}$

Calculate $Z_\lambda^1, Z_\lambda^2, \mathbf{D}_I, \mathbf{D}_{IJ}, \mathbf{D}_J$.

Calculate $P_1, P_2, \ldots, P_8$.

$\nu_{I1}^i = -P_8^{-1}(\mathbf{1}_{m_I} + I_\kappa^\top P_7^{-1}[P_2 P_4^{-1} I_\omega P_5^{-1} \mathbf{1}_{m_I^i} + P_2 P_6 Z_\lambda^2 + Z_\lambda^1])$

$\kappa_I^{i\text{vec}} = -P_7^{-1}(P_2 P_4^{-1} I_\omega P_5^{-1} \mathbf{1}_{m_I^i} + I_\kappa \nu_{I1}^i + P_2 P_6 Z_\lambda^2 + Z_\lambda^1)$

$\eta_I^{i\text{vec}} = -P_3^{-1} I_\kappa \nu_{I1}^i$

$\omega_I^{i\text{vec}} = P_6 P_2^\top \kappa_I^{i\text{vec}} + P_6 Z_\lambda^2 + P_4^{-1} I_\omega P_5^{-1} \mathbf{1}_{m_I^i}$

$\lambda_I^{i\text{vec}} = ((A_{II}^\top A_{II})^\top \circ (\eta_I^i S_I^i \eta_I^{i\top} + (Z_I^i - \kappa_I^i) F_I^i (Z_I^i - \kappa_I^i)^\top))^{-1} \times \text{diag}^{-1}([(\kappa_I^i - Z_I^i) F_I^i \kappa_I^{i\top} + \eta_I^i S_I^i \eta_I^{i\top}] A_{II}^\top A_{II} + (\kappa_I^i - Z_I^i) Q_I^i \omega_I^{i\top} A_I^{\text{inei}\top} A_{II})$

$\lambda_I^{i+} = \text{diag}(\lambda_I^{i\text{vec}})$

until $|\lambda_I^i - \lambda_I^{i+}| \leq \epsilon_I^i$

return $\kappa_I^i = I_\kappa^\top \circ (\mathbf{1}_{m_I} \otimes \kappa_I^{i\text{vec}\top})$, $\eta_I^i = I_\kappa^\top \circ (\mathbf{1}_{m_I} \otimes \eta_I^{i\text{vec}\top})$, $\omega_I^i = I_\omega^\top \circ (\mathbf{1}_{m_I^i} \otimes \omega_I^{i\text{vec}\top})$ and $\lambda_I^i = \lambda_I^{i+}$

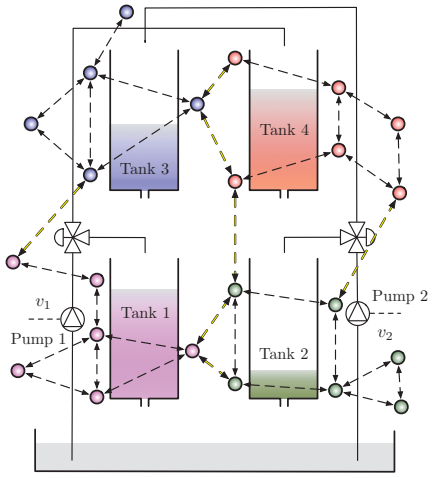

Fig. 1. The 4-tanks system monitored by a sensor network example. Each tank is monitored by a set of sensors (having the same colour as the corresponding tank level). The communication network is represented by dotted arrows (black arrows for local communication, yellow arrows for communication with sensors monitoring neighboring subsystems.)

V. Simulation Results

We consider the four-tanks system introduced in [18] (see Fig. 1 for an example), we linearize and discretize it with a sampling time $T_s = 1$s so to obtain the linear system described by the matrix:

$$A = \begin{bmatrix} 0.9430 & -0.0031 & 0.0262 & -0.0118 \\ -0.0036 & 0.9579 & 0.0121 & 0.0213 \\ -0.0025 & -0.0233 & 0.9500 & -0.0084 \\ -0.0153 & -0.0010 & -0.0053 & 0.9629 \end{bmatrix}.$$

The system can be decomposed in 4 interconnected subsystems and dynamics of each subsystem Σ_I is denoted by $A_{II} = A(I,I), A_{IJ} = A(I,J)$ with $I, J = 1, 2, 3, 4$. We assume the modeling uncertainty ω_I has zero mean value and the system is initialized at $[4.81; 4.70; 1.0; 1.0]^\top$. We compare the following prediction methods:

E_{CKF}: Centralized Kalman one-step ahead predictor.

E_{DKF}: Partition-based method in [6], modified averaging the measurements from different sensors in each subsystem.

E_{p1}: The proposed Pareto-optimal predictor.

The centralized Kalman filter is considered as a benchmark. A sensor network is considered, composed of 40 sensors randomly distributed over a square area with size 40. Each subsystem Σ_I is monitored by 10 sensors. The graph $\mathcal{G}$ is acquired under the rule that two nodes are connected if their relative distance is less than $1.7\sqrt{17}$. We compare the performance of the considered methods in different scenarios with different disturbance and measurement noises. For each noise scenario, we run the experiment 80 times with a random network topology. The evaluated performance metric, denoted as MSE, is the averaged mean square prediction error value over $N_{exp} = 80$ experiments over $N = 40$ nodes, with $T_{max} = 200$s:

$$\text{MSE} := \frac{\sum_{exp=1}^{N_{exp}} \sum_{i=1}^{N} \sum_{t=1}^{T_{max}} \text{MSE}_i(t, exp)}{N_{exp} \cdot N \cdot T_{max}}.$$

Fig. 2 and 3 show MSE values for all methods with different measurement noise standard deviation. The proposed method always has better performance than the other partition-based estimation approach. Moreover, the performance of the proposed method is closer to that of the centralized Kalman predictor when the disturbance standard deviation is low.

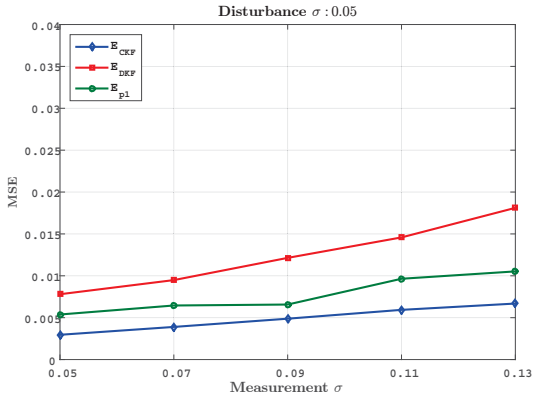

Fig. 2. MSE of 3 prediction methods with different measurement noise standard deviation. Disturbance standard deviation is 0.05

VI. Concluding Remarks

In this paper, a novel distributed partition-based prediction method using sensor networks is proposed for the monitoring

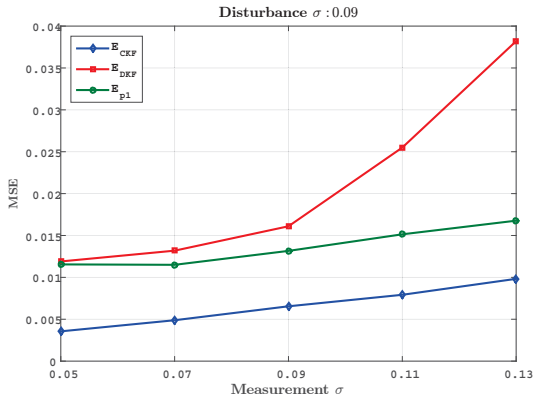

Fig. 3. MSE of 3 prediction methods with different measurement noise standard deviation. Disturbance standard deviation is 0.09

of interconnected stochastic systems. The prediction weights are optimized by each sensor to minimize the bias and the variance of the prediction error at each time step.

REFERENCES

[1] U. A. Khan and J. M. Moura, "Distributing the kalman filter for large-scale systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4919–4935, 2008.

[2] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Consensus based overlapping decentralized estimation with missing observations and communication faults," *Automatica*, vol. 45, no. 6, pp. 1397–1406, 2009.

[3] M. Farina, G. Ferrari-Trecate, and R. Scattolini, "Moving-horizon partition-based state estimation of large-scale systems," *Automatica*, vol. 46, no. 5, pp. 910–918, 2010.

[4] F. Boem, R. Ferrari, T. Parisini, and M. Polycarpou, "Distributed fault detection and isolation of continuous-time nonlinear systems," *European Journal of Control*, vol. 5-6, pp. 603–620, 2011.

[5] S. Riverso, D. Rubini, and G. Ferrari-Trecate, "Distributed bounded-error state estimation for partitioned systems based on practical robust positive invariance," in *European Control Conf.*, July 2013, pp. 2633–2638.

[6] M. Farina and R. Carli, "Plug and play partition-based state estimation based on kalman filter," in *54th IEEE Conf. on Decision and Control*, Dec 2015, pp. 3155–3160.

[7] R. Olfati-Saber, "Kalman-consensus filter : Optimality, stability, and performance," in *Decision and Control, 48th IEEE Conference on, and 28th Chinese Control Conference*, 2009, pp. 7036–7042.

[8] ——, "Distributed Kalman filtering for sensor networks," in *Decision and Control, 46th IEEE Conference on*, 2007, pp. 5492–5498.

[9] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 622 –633, 2008.

[10] J. Liang, Z. Wang, and X. Liu, "Distributed state estimation for discrete-time sensor networks with randomly varying nonlinearities and missing measurements," *IEEE Transactions on Neural Networks*, vol. 22, no. 3, pp. 486–496, March 2011.

[11] D. Wang, Z. Wang, G. Li, and W. Wang, "Distributed filtering for switched nonlinear positive systems with missing measurements over sensor networks," *IEEE Sensors Journal*, vol. 16, no. 12, pp. 4940–4948, June 2016.

[12] G. Battistelli and L. Chisci, "Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.

[13] F. Boem, Y. Xu, C. Fischione, and T. Parisini, "A distributed estimation method for sensor networks based on Pareto optimization," in *51th IEEE Conf. on Decision and Control*, 2012, pp. 775–781.

[14] ——, "Distributed fault detection using sensor networks and Pareto estimation," in *European Control Conf.*, 2013, pp. 932–937.

[15] ——, "A distributed Pareto-optimal dynamic estimation method," in *European Control Conf.*, 2015, pp. 3678–3685.

[16] P. Alriksson and A. Rantzer, "Distributed Kalman filtering using weighted averaging," in *17th Int. Symposium on Mathematical Theory of Networks and Systems*, 2006, pp. 1–60.

[17] A. Speranzon, C. Fischione, K. Johansson, and A. Sangiovanni-Vincentelli, "A distributed minimum variance estimator for sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 609–621, 2008.

[18] K. Johansson, "The quadruple-tank process: a multivariable laboratory process with an adjustable zero," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 456–465, May 2000.

VII. APPENDIX

KKT conditions

$$(\kappa_I^i + \eta_I^i)\mathbf{1}_{N_I^i m_I} - \mathbf{1}_{m_I} = \mathbf{0}_{m_I}, \tag{23}$$

$$\omega_I^i \mathbf{1}_{p_I^i} - \mathbf{1}_{m_I^i} = \mathbf{0}_{m_I^i}, \tag{24}$$

$$\lambda_I^i \mathbf{1}_{m_I} - \mathbf{1}_{m_I} \leq \mathbf{0}_{m_I} \tag{25}$$

$$-\lambda_I^i \mathbf{1}_{m_I} \leq \mathbf{0}_{m_I}, \tag{26}$$

$$\xi_{I1}^{i\top}(\lambda_I^i \mathbf{1}_{m_I} - \mathbf{1}_{m_I}) = 0 \qquad \xi_{I1}^i \geq 0, \tag{27}$$

$$\xi_{I2}^{i\top}(-\lambda_I^i \mathbf{1}_{m_I}) = 0 \qquad \xi_{I2}^i \geq 0, \tag{28}$$

$$-\kappa_I^{i\top} \mathbf{1}_{m_I} \leq \mathbf{0}_{N_I^i m_I^i}, \tag{29}$$

$$\kappa_I^i \mathbf{1}_{N_I^i m_I} - \frac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I} + \epsilon_{I1}^i \leq \mathbf{0}_{m_I}, \tag{30}$$

$$\xi_{I3}^{i\top}(-\kappa_I^{i\top} \mathbf{1}_{m_I}) = 0 \qquad \xi_{I3}^i \geq 0, \tag{31}$$

$$\xi_{I4}^{i\top}(\kappa_I^i \mathbf{1}_{N_I^i m_I} - \frac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I} + \epsilon_{I1}^i) = 0 \qquad \xi_{I4}^i \geq 0, \tag{32}$$

$$-\omega_I^{i\top} \mathbf{1}_{m_I^i} \leq \mathbf{0}_{p_I^i} \tag{33}$$

$$\omega_I^i \mathbf{1}_{p_I^i} - \frac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I^i} + \epsilon_{I2}^i \leq \mathbf{0}_{m_I^i} \tag{34}$$

$$\xi_{I5}^{i\top}(-\omega_I^{i\top} \mathbf{1}_{m_I^i}) = 0 \qquad \xi_{I5}^i \geq 0, \tag{35}$$

$$\xi_{I6}^{i\top}(\omega_I^i \mathbf{1}_{p_I^i} - \frac{1}{\|A_I^{Eg}\|_\infty}\mathbf{1}_{m_I^i} + \epsilon_{I2}^i) = 0 \qquad \xi_{I6}^i \geq 0, \tag{36}$$

$$\begin{aligned} &2[F_I^i(\rho_I^i)((I-\lambda_I^i)\kappa_I^i + \lambda_I^i Z_I^i)^\top A_{II}^\top A_{II}(I-\lambda_I^i) \\ &+ Q_I^i(\rho_I^i)\omega_I^{i\top} A_I^{\text{inei}\top} A_{II}(I-\lambda_I^i)] \circ I_\kappa \\ &+ [\mathbf{1}_{N_I^i m_I}\nu_{I1}^{i\top} + \mathbf{1}_{N_I^i m_I}\xi_{I4}^{i\top} - \xi_{I3}^i \mathbf{1}_{m_I}^\top] \circ I_\kappa = 0, \end{aligned} \tag{37}$$

$$\begin{aligned} &2[S_I^i(\rho_I^i)\eta_I^{i\top}(I-\lambda_I^i)^\top A_{II}^\top A_{II}(I-\lambda_I^i)] \circ I_\kappa \\ &+ [\mathbf{1}_{N_I^i m_I}\nu_{I1}^{i\top}] \circ I_\kappa = 0, \end{aligned} \tag{38}$$

$$\begin{aligned} &2[[\eta_I^i S_I^i(\rho_I^i)\eta_I^{i\top} + (Z_I^i - \kappa_I^i)F_I^i(\rho_I^i)(Z_I^i - \kappa_I^i)^\top]\lambda_I^i \\ &\times A_{II}^\top A_{II} + [(Z_I^i - \kappa_I^i)F_I^i(\rho_I^i)\kappa_I^{i\top} - \eta_I^i S_I^i(\rho_I^i)\eta_I^{i\top}] \\ &\times A_{II}^\top A_{II} + (Z_I^i - \kappa_I^i)Q_I^i(\rho_I^i)\omega_I^{i\top} A_I^{\text{inei}\top} A_{II}] \circ I_\lambda \\ &+ [\mathbf{1}_{m_I}\xi_{I1}^{i\top} - \mathbf{1}_{m_I}\xi_{I2}^{i\top}] \circ I_\lambda = 0, \end{aligned} \tag{39}$$

$$\begin{aligned} &2[Q_I^{i\top}(\rho_I^i)((I-\lambda_I^i)\kappa_I^i + \lambda_I^i Z_I^i)^\top A_{II}^\top A_I^{\text{inei}} \\ &+ F_I^{\text{inei}}(\rho_I^i)\omega_I^{i\top} A_I^{\text{inei}\top} A_I^{\text{inei}}] \circ I_\omega \\ &+ [\mathbf{1}_{p_I^i}\nu_{I2}^{i\top} + \mathbf{1}_{p_I^i}\xi_{I6}^{i\top} - \xi_{I5}^i \mathbf{1}_{m_I^i}^\top] \circ I_\omega = 0 \end{aligned} \tag{40}$$

where ϵ_{I1}^i and ϵ_{I2}^i are small positive constants.