IET Control Theory & Applications

Research Article

Power scheduling for Kalman filtering over lossy wireless sensor networks

ISSN 1751-8644
Received on 7th August 2016
Revised 19th October 2016
Accepted on 17th November 2016
E-First on 13th December 2016
doi: 10.1049/iet-cta.2016.1047
www.ietdl.org

Jie Chen[1], Gang Wang[1,2], Jian Sun[1] ✉

[1]*School of Automation, Beijing Institute of Technology, Beijing 100081, People's Republic of China*
[2]*ECE Department, University of Minnesota, Minneapolis, MN 55455, USA*
✉ *E-mail: sunjian@bit.edu.cn*

Abstract: With the goal of monitoring physical processes, a wireless sensor network (WSN) is often deployed along with a fusion center to estimate the state of general linear stochastic systems. As WSNs comprise a large number of low-cost, battery-driven sensor nodes with limited transmission bandwidth, conservation of transmission resources (power and bandwidth) is of paramount importance. In this context, the present study considers power scheduling for Kalman filtering (KF) using scalar messages exchanged over wireless sensor links, where random measurement packet drops are possible. Each sensor node sequentially decides whether a high or low transmission power is needed to communicate its scalar observations based on a rule that promotes power scheduling with minimal impact on the state estimator's mean-squared error. Assuming approximately Gaussian state predictors, the minimum mean-squared error optimal power schedule is developed for KF that also accounts for dropped data packets. Leveraging statistical convergence characteristics of the estimation error covariance matrix, both sufficient and necessary conditions are established that guarantee the stability of the resultant KF estimator.

1 Introduction

With the advent of microsensor and wireless communication technologies, wireless sensor networks (WSNs) have found diverse applications, including surveillance, intelligent transportation systems, health care, environmental tracking, disaster prevention and recovery, as well as monitoring of electric power grids [1–4]. Typical WSN attributes are battery-driven sensors, limited computational and communication capabilities under stringent bandwidth constraints [3]. These limitations ineluctably bring challenges to estimation and control tasks performed using WSNs. Therefore, it is critical to investigate how to conserve transmission power and bandwidth while achieving a prescribed estimation performance.

Towards this end, recurring attention has been paid to state estimation at the fusion center (FC) under stringent constraints on communication resources (energy and bandwidth); see, for example, [3–18] and references therein. To save transmission power and bandwidth, various methods relying on measurement quantisation/censoring, dimensionality reduction, multi-rate transmission, and scheduling were pursued in [3–5, 8, 10–17, 19–21].

In scheduling-based approaches, optimal policies were introduced in [11] for a class of scalar linear stochastic systems to minimise the terminal state estimation error variance over a fixed time horizon T, in which only $p < T$ measurements can be transmitted to the FC. In practice, most commercially available sensor nodes nowadays have multiple transmission-power levels [8]. Clearly, high transmission power leads to reliable message exchanges while low transmission power may cause data packets to drop [12]. The results in [11] were recently extended to a special class of high-order linear systems, where not only energy constraints and data packet drops were taken into account, but also sensors with limited or sufficient computational capacity were considered [13]. Under certain conditions, the optimal schedulers in [11] amount to distributing the p measurement transmissions along the last p time instants over the horizon T. The aforementioned optimal schedulers are deterministic, and are obtained offline. Nevertheless, their state estimation error covariance matrix increases drastically for unstable systems in the first $T - p$ time instants because no measurements are transmitted to the FC for updating the covariance prediction; see also [12, 17] for further generalisations.

To cope with this instability, online schedulers were developed in [14–16]. Specifically, a so-termed send-on-delta strategy was adopted in [14] to reduce sensor data traffic by transmitting sensor data only if their values change more than a prescribed threshold. However, neither this threshold was analytically selected to improve the estimation performance, nor stability or performance analyses were given for the resultant modified Kalman filtering (KF). Innovation-based measurement schedulers were also constructed in [15, 16] by quantifying the 'importance' of every measurement using the normalised measurement innovations. The key idea is that only 'sufficiently important' measurements are transmitted to the FC for updating the predicted estimate and its covariance, and when the transmission does not occur, the available information based on a threshold selected by the scheduler is utilised. Stability analyses of the KF with these stochastic schedulers were reported in [15]. Yet, only necessary conditions ensuring convergence of the expected state estimation error covariance matrix were established for systems with observation matrix having full row rank.

Inspired by these works and building on our precursor in [22], this paper considerably broadens the scope of [3, 12, 15], where the power scheduler depends on the time-horizon T and the state error covariance increases during the first $T - p$ time instants. In comparison, the main contribution of this work is two-fold and can be summarised as follows.

(a) We consider power scheduling for KF-based state estimation of general linear stochastic systems. Data packet drops are accounted for in developing an innovation-based power scheduler, and the corresponding minimum mean-squared error (MMSE) state estimator.
(b) We investigate statistical convergence of the state estimation error covariance matrix, and establish both sufficient and necessary conditions for convergence of the averaged estimation error covariance.

Error estimator creatively devised in [15] are extended to the case where different thresholds are assigned to different components of a measurement vector. On the other hand, both the sufficient condition and the necessary condition are derived for

general linear dynamic systems by analysing convergence properties of an axillary function.

In a nutshell, the present contribution generalises the results of [3, 15, 23–25] of state estimation using WSNs both in the estimation setup, as well as in analysing the estimator's stability.

Notation: Boldface symbols denote the multivariate quantities such as vectors (lowercase) and matrices (uppercase); $Q(x) := (1/\sqrt{2\pi})\int_x^\infty \exp(-t^2/2)\,\mathrm{d}t$ is the tail probability of the standardised normal probability density function (pdf); $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the normal pdf with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For random vectors $\mathbf{x}$ and $\mathbf{y}$, $E[\mathbf{x}]$ denotes the expectation of $\mathbf{x}$, and $\mathbf{x}|\mathbf{y}$ is the conditional random vector when $\mathbf{y}$ is given. Furthermore, $(\cdot)^{\mathrm{T}}$ stands for transposition; $\mathbf{P} > \mathbf{0}\,(\geq \mathbf{0})$ for the positive (semi-)definite matrix, and $\mathrm{diag}\{l_1, l_2, \ldots, l_m\}$ for the diagonal matrix with diagonal elements $l_1, l_2, \ldots, l_m$; $\mathbf{I}_n$ denotes the $n \times n$ identity matrix and $\mathbf{0}$ the all-zero matrix of appropriate dimensions; and $\otimes$ is reserved for the Kronecker product of two matrices. The mean-square stability of the filter, defined as, $\sup_{k \in N} \mathbb{E}[\mathbf{P}_k] < \infty$, implies there always exists a positive definite matrix $\bar{\mathbf{P}}$ such that $\mathbf{P}_k \leq \bar{\mathbf{P}}$ for all $k \in N$ [23], where the mathematical expectation is taken over both the random power scheduling process and the random packet dropping process. For positive (semi-)definite matrices $\mathbf{P}$ and $\mathbf{Q}$, the matrix inequality $\mathbf{P} \geq \mathbf{Q}$ means matrix $\mathbf{P} - \mathbf{Q}$ is positive semi-definite; and likewise, for $\mathbf{P} \leq \mathbf{Q}, \mathbf{P} > \mathbf{Q}$ and $\mathbf{P} < \mathbf{Q}$.

2 Problem statement and preliminaries

Consider a WSN with m sensor nodes $\{\mathcal{S}_i\}_{i=1}^m$ deployed to estimate the state $\mathbf{x}_k \in \mathbb{R}^n$, obeying the recursion

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \boldsymbol{\omega}_k \tag{1}$$

where k denotes time index, $\mathbf{A} \in \mathbb{R}^{n \times n}$ the state transition matrix, and $\boldsymbol{\omega}_k \in \mathbb{R}^n$ the process noise, assumed to be a zero-mean white Gaussian with covariance matrix $\mathbf{Q} \geq \mathbf{0}$. The initial state $\mathbf{x}_0$ is also assumed Gaussian distributed with mean $\hat{\boldsymbol{x}}_0$ and covariance matrix $\mathbf{P}_0 > \mathbf{0}$.

Each sensor indexed by i receives scalar observations $\{y_k^i\}$ adhering to a linear measurement equation

$$y_k^i = \mathbf{c}_i^{\mathrm{T}}\mathbf{x}_k + \nu_k^i \tag{2}$$

where $\mathbf{c}_i \in \mathbb{R}^n$ is the regression vector, and ν_k^i is a temporally and spatially white zero-mean Gaussian noise with variance r_i. Collecting all sensor observations leads to the vector-matrix counterpart of (2), namely $\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \boldsymbol{\nu}_k$ with $\mathbf{y}_k := [\mathbf{y}_k^1 \cdots \mathbf{y}_k^m]^{\mathrm{T}}$, $\mathbf{C} := [\mathbf{c}_1 \cdots \mathbf{c}_m]^{\mathrm{T}}$ and $\boldsymbol{\nu}_k := [\nu_k^1 \cdots \nu_k^m]^{\mathrm{T}}$. It is further posited that the random vectors $\boldsymbol{\omega}_k, \boldsymbol{\nu}_k, \mathbf{x}_0$ are mutually independent, and that the following standard assumption holds.

(As1) *Pairs* $(\mathbf{A}, \mathbf{Q}^{1/2})$ *and* $(\mathbf{C}, \mathbf{A})$ *are controllable and observable, respectively.*

The FC is a designated node, performs the task of estimating $\boldsymbol{x}_k$ from noisy measurements y_k transmitted by the sensors over wireless sensor links, where packet drops are possible. Only transmissions between the FC and the sensors are allowed, implying no inter-sensor transmissions, and leading to lower communication costs. The FC is assumed able to feed information back to the sensors. As in [4], a round-robin, slotted-time sensor schedule is envisioned: A sampling interval (time between k and $k+1$) is partitioned into m time slots (one per sensor) such that sensor $\mathcal{S}_i$ transmits at the ith time slot T_i.

Clearly, a high transmission-power improves the reliability of measurement exchanges, while a low transmission power may cause packet drops during sensor-to-FC communications. This is reasonable and it is motivated by two facts: most available sensors have multiple transmission-power levels to choose from [8], and higher transmission-power leads to a higher signal-to-noise ratio (SNR) at the FC. In turn, higher SNR implies higher packet arrival rate [26]. Therefore, once a communication failure occurs, the entire packet of measurements will be dropped.

For simplicity, the present work considers that each sensor has only two transmission-power levels [12, 13]. However, all results derived can be easily generalised to multiple transmission-power levels. Specifically, with a high transmission power Δ, the measurement packet will be assumed successfully delivered to the FC. If on the other hand a low transmission power δ is employed, the measurement packet will be received correctly at the FC only with probability $\beta \in (0, 1)$. Although in a different estimation setup, similar assumptions were adopted by Shi and Xie [13].

Per slot i per time instant k, binary random variable $\gamma_k^i \in \{0, 1\}$, represents whether transmission power δ or Δ is utilised for transmitting y_k^i to the FC. Another binary random variable $\beta_k^i \in \{1, 0\}$ indicates whether y_k^i is successfully received at the FC or not. Throughout this paper, we postulate that the values of γ_k^i and β_k^i are known at the FC. This is reasonable if one employs the time-stamp technique [25, 27]. For future reference, define the following sets capturing all historical information up to time k, slot i, including all received measurements, along with the power levels transmitted and success or failure of reception $\mathcal{I}_k^i := \{\gamma_1^1 y_1^1, (1-\gamma_1^1)\beta_1^1 y_1^1, \gamma_1^1, (1-\gamma_1^1)\beta_1^1, \gamma_1^2 y_1^2, (1-\gamma_1^2)\beta_1^2 y_1^2, \gamma_1^2, (1-\gamma_1^2)\beta_1^2, \ldots, \gamma_k^i y_k^i, (1-\gamma_k^i)\beta_k^i y_k^i, \gamma_k^i, (1-\gamma_k^i)\beta_k^i\}$, and $\mathcal{G}_k^i := \{\mathcal{I}_k^{i-1}, y_k^i\}$, $i = 1, 2, \ldots, m$. To understand the two sets, let us consider, for example, a system having two sensors with their observations at time k denoted by $\{y_k^1, y_k^2\}$. Thus, every time interval from t to $t+1$ is spitted into two time slots $i = 1$ and $i = 2$. At slot $i = 2$ of time k, the set $\mathcal{I}_k^1$ consisting of all information up to slot 2 time k can be represented by $\mathcal{I}_k^1$ that up to slot 1 time k plus the new information $\{\gamma_k^2 y_k^2, (1-\gamma_k^2)\beta_k^2 y_k^2, \gamma_k^2\}$ arriving within the new slot. It is not clear whether y_k^2 is included in $\mathcal{I}_k^2$ depending on whether y_k^2 reaches successfully the estimator or not. The set $\mathcal{G}_k^2$ is thus defined to include also y_k^2. Then at slot 1 of $k+1$, $\mathcal{I}_{k+1}^1$ can be written as $\mathcal{I}_k^2$ all information up to the last slot plus $\{\gamma_{k+1}^1 y_{k+1}^1, (1-\gamma_{k+1}^1)\beta_{k+1}^1 y_{k+1}^1, \gamma_{k+1}^1\}$ the new information reaching within the new slot.

Thus per time instant k, let $\hat{\mathbf{x}}_{k|k}^i$ denote the MMSE estimate of $\mathbf{x}_k$ at the FC based upon all available information at the end of slot i; and, likewise $\mathbf{P}_{k|k}^i$ its state estimation error covariance matrix; that is

$$\hat{\mathbf{x}}_{k|k}^i = \mathbb{E}\left[\mathbf{x}_k | \mathcal{I}_k^i\right] \tag{3}$$

$$\mathbf{P}_{k|k}^i = \mathbb{E}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)^{\mathrm{T}} | \mathcal{I}_k^i\right]. \tag{4}$$

Instrumental to the ensuing derivations are the so termed predicted estimate and predicted estimation error covariance matrix

$$\hat{\mathbf{x}}_{k|k-1} = \mathbb{E}\left[\mathbf{x}_k | \mathcal{I}_{k-1}^m\right] \tag{5}$$

$$\mathbf{P}_{k|k-1} = \mathbb{E}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^m)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}^m)^{\mathrm{T}} | \mathcal{I}_{k-1}^m\right]. \tag{6}$$

The measurement innovation is well acknowledged to represent new information of the current measurement that cannot be predicted by past measurements. Intuitively speaking, a large innovation means that the current measurement is quite different than the predicted measurement and therefore contains useful information to update the estimates. Thus, measurements with sizable innovation magnitude are deemed as 'informative', and the vice versa for measurements with small innovation magnitude. Our innovation-based power scheduling policy compares the

normalised measurement innovation magnitude with a given threshold to quantify the 'importance' of every measurement, and then uses a high (or low) transmission power to send the 'important' (or 'less important') measurement.

To be specific, let $\{\eta_i\}_{i=1}^m$ be given fixed thresholds for power scheduling (one per sensor), which can be determined based on the specific requirement on the power saving per sensor. At slot i of time instant k, assuming that the FC has broadcast the estimate $\hat{\mathbf{x}}_{k|k}^{i-1}$ and its estimation error covariance $\mathbf{P}_{k|k}^{i-1}$ to sensor $\mathcal{S}_i$, then sensor $\mathcal{S}_i$ uses them to compute a normalised innovation of the current local observation y_k^i as

$$\epsilon_k^i := \left(y_k^i - \mathbf{c}_i^{\mathrm{T}}\hat{\mathbf{x}}_{k|k}^{i-1}\right)/\sqrt{\mathbf{c}_i^{\mathrm{T}}\mathbf{P}_{k|k}^{i-1}\mathbf{c}_i + r_i}. \tag{7}$$

If $|\epsilon_k^i|$ is greater than the given threshold η_k^i, y_k^i is deemed informative enough and will be transmitted to the FC with high power Δ, and thus $\gamma_k^i = 1$; otherwise, δ will be used and $\gamma_k^i = 0$. The FC will correspondingly resort to different rules to update $(\hat{\mathbf{x}}_{k|k}^{i-1}, \mathbf{P}_{k|k}^{i-1})$ to obtain $(\hat{\mathbf{x}}_{k|k}^{i}, \mathbf{P}_{k|k}^{i})$, which is further elaborated in the next section. Information broadcast from the FC can reach successfully all sensors since the FC has enough power and wireless sensors also consume much less power for receiving than sending packets [28].

3 Kalman filtering with power scheduling

Algorithms for power scheduling and estimation steps that are amenable to WSN implementation are elaborated in this section. Then the developed power scheduler and the corresponding MMSE estimator are, respectively, tabulated as Algorithms 1 and 2).

Algorithm 1 (Scheduling and transmission per sensor i): Sensor $\mathcal{S}_i$ at time k

Require: $\hat{\mathbf{x}}_{k|k}^{i-1}$ and $\mathbf{P}_{k|k}^{i-1}$

Ensure: $\tilde{y}_k^i$

S1: Compute $\tilde{y}_k^i := y_k^i - \mathbf{c}_i^{\mathrm{T}}\hat{\mathbf{x}}_{k|k}^{i-1}$ and $\sigma_k^i := \sqrt{\mathbf{c}_i^{\mathrm{T}}\mathbf{P}_{k|k}^{i-1}\mathbf{c}_i + r_i}$

S2: Compute $\epsilon_k^i := \tilde{y}_k^i/\sigma_k^i$

S3: Using prescribed η_i, find

$$\gamma_k^i = \begin{cases} 1 & \text{if } |\epsilon_k^i| > \eta_i \\ 0 & \text{otherwise} \end{cases}$$

S4: If $\gamma_k^i = 1$ then

$\mathcal{S}_i$: Transmit $\tilde{y}_k^i$ to FC with power Δ

else

$\mathcal{S}_i$: Transmit $\tilde{y}_k^i$ to FC with power δ

Algorithm 2 (Reception and estimation at the FC): **Require:** $\hat{\mathbf{x}}_{0|0}$ and $\mathbf{P}_{0|0}$

S1: **for** $k = 0$ to ∞ **do**

S2: Compute

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}\hat{\mathbf{x}}_{k-1|k-1}$$
$$\mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$$

S3: **for** $i = 1$ to m **do**

S4: Receive y_k^i and $\{\gamma_k^i, \beta_k^i\}$ from sensor i

S5: Compute

$$l(\beta_k^i) := \beta_k^i + (1-\beta_k^i)\sqrt{\frac{2}{\pi}} \times \frac{\eta_i \exp(-\eta_i^2/2)}{1 - 2Q(\eta_i)}$$
$$s(\gamma_k^i, \beta_k^i) := \gamma_k^i + (1-\gamma_k^i)\beta_k^i$$
$$t(\gamma_k^i, \beta_k^i) := \gamma_k^i + (1-\gamma_k^i)l(\beta_k^i)$$
$$\mathbf{k}_k^i := \mathbf{P}_{k|k}^{i-1}\mathbf{c}_i/(\mathbf{c}_i^{\mathrm{T}}\mathbf{P}_{k|k}^{i-1}\mathbf{c}_i + r_i)$$
$$\hat{\mathbf{x}}_{k|k}^i = \hat{\mathbf{x}}_{k|k}^{i-1} + s(\gamma_k^i, \beta_k^i)\mathbf{k}_k^i(y_k^i - \mathbf{c}_i^{\mathrm{T}}\hat{\mathbf{x}}_{k|k}^{i-1})$$
$$\mathbf{P}_{k|k}^i = \mathbf{P}_{k|k}^{i-1} - t(\gamma_k^i, \beta_k^i)\mathbf{k}_k^i\mathbf{c}_i^{\mathrm{T}}\mathbf{P}_{k|k}^{i-1}$$

S6: **end for**

S7: $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k}^m$, $\mathbf{P}_{k|k} = \mathbf{P}_{k|k}^m$

S8: **end for**

Remark 1: Channel propagation effects between the sensor and the FC have not been considered here. As pointed out in [8], most of the existing works in distributed estimation assumed error-free reception between the FC and the sensors. Nonetheless, focusing on the sensor-to-FC transmissions of y_k^i per time slot, suppose that the corresponding channel has gain g_k^i. Postulating that the fading is relatively slow, the phase of the complex channel can be estimated and thus compensated for at the receiver side, so that g_k^i stands for the real-valued envelop of the complex channel gain [29]. Further, suppose that the channel gain remains invariant over the time slot to send y_k^i. Then the FC receives a scaled version of y_k^i corrupted with the additional noise which is independent of the measurement noise. For simplicity, let the channel noise n_k^i be zero-mean Gaussian white with variance $\sigma_{n_i}^2$ and, $\{n_k^i\}$, $\{n_k^j\}$ be uncorrelated for $i \neq j$.

In light of the round-robin, time-slotted transmission policy and (2), one further arrives at $z_k^i = g_k^i\mathbf{c}_i^{\mathrm{T}}\mathbf{x}_k + g_k^i\nu_k^i + n_k^i$. Upon defining $\bar{\nu}_k^i = g_k^i\nu_k^i + n_k^i$, it follows that

$$z_k^i = g_k^i\mathbf{c}_i^{\mathrm{T}}\mathbf{x}_k + \bar{\nu}_k^i. \tag{8}$$

If the gain g_k^i is invariant over time instant k, then by letting $\bar{\mathbf{c}}_k^i = g_k^i\mathbf{c}_i$, this model reduces to (2); see also [29], where KF with faded observations was reported along with stability analysis. Hence results in this paper can be generalised to account for fading effects along the line of [29].

Proposition 1: If the conditional pdf of $\mathbf{x}_k$ given $\mathcal{F}_{k-1}^{i-1}$ is approximately Gaussian, i.e. the pdf $f\left(\mathbf{x}_k|\mathcal{F}_k^{i-1}\right) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{i-1}, \mathbf{P}_{k|k}^{i-1})$, then $\hat{\mathbf{x}}_{k|k}^i$ in Algorithm 2 yields the MMSE estimator.

Remark 2: Since the pdf $f\left(\mathbf{x}_k|\mathcal{F}_k^{i-1}\right)$ is in general non-Gaussian, computationally expensive numerical integrations and memory intensive propagation of the posterior pdf are required for computing the exact MMSE estimate. This motivates approximating this pdf with a Gaussian one, which is a customary simplification adopted in non-linear filtering [30] and the KF with quantised innovations [3]; see, also, [10, 15] and references therein.

Proof of Proposition 1:: Suppose we already have an MMSE estimator $\hat{\mathbf{x}}_{k|k}^{i-1}$, that is, $\hat{\mathbf{x}}_{k|k}^{i-1} = \mathbb{E}\left[\mathbf{x}_k|\mathcal{F}_k^{i-1}\right]$ and $\mathbf{P}_{k|k}^{i-1} = \mathbb{E}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{i-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{i-1})^{\mathrm{T}}|\mathcal{F}_k^{i-1}]$. We prove the proposition by conditioning on whether the measurement is received by the FC. Specifically, when the new measurement y_k^i is present at the FC, that is, the case $\gamma_k^i = 1$ or the case $\gamma_k^i = 0$ and $\beta_k^i = 1$, one can easily verify that

$$\hat{\mathbf{x}}_{k|k}^i = E\left[\mathbf{x}_k|\mathcal{F}_k^{i-1}, y_k^i\right] = \hat{\mathbf{x}}_{k|k}^{i-1} + \mathbf{k}_k^i\left(y_k^i - \mathbf{c}_i^{\mathrm{T}}\hat{\mathbf{x}}_{k|k}^{i-1}\right)$$

and likewise,

$$\begin{aligned}\mathbf{P}^i_{k|k} &= E\left[(\mathbf{x}_k - \hat{\mathbf{x}}^i_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}^i_{k|k})^{\mathrm{T}} | \mathcal{F}^{i-1}_k, y^i_k\right] \\ &= \mathbf{P}^{i-1}_{k|k} - \mathbf{P}^{i-1}_{k|k}\mathbf{c}_i\left(\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\mathbf{c}_i + r_i\right)^{-1}\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}.\end{aligned}$$

When the estimator does not receive the new measurement y^i_k, that is, $\gamma^i_k = 0$ and $\beta^i_k = 0$, then it follows that

$$\begin{aligned}\hat{\mathbf{x}}^i_{k|k} &= \mathbb{E}\left[\mathbf{x}_k | \mathcal{F}^{i-1}_k, \gamma^i_k = 0, \beta^i_k = 0\right] \\ &= \mathbb{E}\left[\mathbf{x}_k | \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right] \\ &= \int_{-\eta_i}^{\eta_i} (\hat{\mathbf{x}}^{i-1}_{k|k} + \mathbf{k}^i_k\sigma^i_k\epsilon) f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right) \mathrm{d}\epsilon.\end{aligned} \tag{9}$$

Here, $f_{\mathbf{x}}(x)$ is the pdf of the random variable $\mathbf{x}$; similarly, $f_{\mathbf{x}|\mathbf{y}}(x|y)$ is the pdf of a random variable $\mathbf{x}$ conditioned on variable y. Given $\mathcal{F}^{i-1}_k$, then ϵ^i_k is Gaussian distributed with zero-mean and unit covariance. Thus, the conditional pdf above follows directly from conditional probability theory:

$$f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right) = \begin{cases} \dfrac{f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k\right)}{\Delta P_i}, & \text{if } |\epsilon^i_k| \le \eta_i \\ 0, & \text{otherwise}\end{cases} \tag{10}$$

where $\triangle P_i \triangleq \Pr\left(|\epsilon^i_k| \le \eta_i | \mathcal{F}^{i-1}_k\right) = 1 - 2Q(\eta_i)$. Therefore, (9) becomes

$$\begin{aligned}\hat{\mathbf{x}}^i_{k|k} &= \int_{-\eta_i}^{\eta_i} \left(\hat{\mathbf{x}}^{i-1}_{k|k} + \mathbf{k}^i_k\sigma^i_k\epsilon\right) \frac{f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k\right)}{\triangle P_i} \mathrm{d}\epsilon \\ &= \hat{\mathbf{x}}^{i-1}_{k|k} \int_{-\eta_i}^{\eta_i} \frac{f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k\right)}{\triangle P_i} \mathrm{d}\epsilon + \frac{\mathbf{k}^i_k\sigma^i_k}{\triangle P_i} \int_{-\eta_i}^{\eta_i} \epsilon f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k\right) \mathrm{d}\epsilon \\ &= \hat{\mathbf{x}}^{i-1}_{k|k}\end{aligned} \tag{11}$$

where the first integration equals to 1 and the second becomes 0 because $f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k\right)$ is even over the origin-centred symmetric integration interval.

The covariance $\mathbf{P}^i_{k|k}$ for $\gamma^i_k = 0$ and $\beta^i_k = 0$ case is then

$$\begin{aligned}\mathbf{P}^i_{k|k} &\overset{(a)}{=} \mathbb{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}^i_{k|k}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}^i_{k|k}\right)^{\mathrm{T}} \middle| \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right] \\ &\overset{(b)}{=} \mathbb{E}\left[\left(\mathbf{x}_k - \hat{\mathbf{x}}^{i-1}_{k|k}\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}^{i-1}_{k|k}\right)^{\mathrm{T}} \middle| \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right] \\ &\overset{(c)}{=} \mathbb{E}\Big[\left(\mathbf{x}_k - \hat{\mathbf{x}}^{i-1}_{k|k} - \mathbf{k}^i_k\sigma^i_k\epsilon + \mathbf{k}^i_k\sigma^i_k\epsilon\right) \\ &\quad \times \left(\mathbf{x}_k - \hat{\mathbf{x}}^{i-1}_{k|k} - \mathbf{k}^i_k\sigma^i_k\epsilon + \mathbf{k}^i_k\sigma^i_k\epsilon\right)^{\mathrm{T}} \Big| \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\Big] \\ &\overset{(d)}{=} \mathbb{E}\Big[\left(\left(I_n - \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\right)(\mathbf{x}_k - \hat{\mathbf{x}}^{i-1}_{k|k}) + \mathbf{k}^i_k\nu^i_k + \mathbf{k}^i_k\sigma^i_k\epsilon\right) \\ &\quad \left(\left(\mathbf{I}_n - \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\right)(\mathbf{x}_k - \hat{\mathbf{x}}^{i-1}_{k|k}) + \mathbf{k}^i_k\nu^i_k + \mathbf{k}^i_k\sigma^i_k\epsilon\right)^{\mathrm{T}} \Big| \\ &\quad \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\Big] \\ &\overset{(e)}{=} \left(\mathbf{I}_n - \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\right)\mathbf{P}^{i-1}_{k|k}\left(\mathbf{I}_n - \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\right)^{\mathrm{T}} + \mathbf{k}^i_k r_i (\mathbf{k}^i_k)^{\mathrm{T}} \\ &\quad + \left(\sigma^i_k\right)^2\mathbf{k}^i_k\mathbb{E}\left[\epsilon | \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right]\left(\mathbf{k}^i_k\right)^{\mathrm{T}}\end{aligned} \tag{12}$$

where (b) follows directly from $\hat{\mathbf{x}}^i_{k|k} = \hat{\mathbf{x}}^{i-1}_{k|k}$ when the new measurement component y^i_k is not received by the estimator, which has been proved in (11); and (d) is because $\sigma^i_k\epsilon^i_k = z^i_k = y^i_k - \mathbf{c}_i^{\mathrm{T}}\hat{\mathbf{x}}^{i-1}_{k|k} = \mathbf{c}_i^{\mathrm{T}}(\mathbf{x}_k - \hat{\mathbf{x}}^{i-1}_{k|k}) + \nu^i_k$ in Algorithm 2; and (e) is because ν^i_k is zero-mean Gaussian noise with covariance r_i, or, $\mathbb{E}\left[\mathbf{k}^i_k\nu^i_k(\mathbf{k}^i_k\nu^i_k)^{\mathrm{T}}\right] = \mathbf{k}^i_k r_i (\mathbf{k}^i_k)^{\mathrm{T}}$. Meanwhile, we have

$$\begin{aligned}\mathbb{E}\left[\epsilon^2 | \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right] &= \int_{-\eta_i}^{\eta_i} \epsilon^2 f_{\epsilon^i_k}\left(\epsilon | \mathcal{F}^{i-1}_k, |\epsilon^i_k| \le \eta_i\right) \mathrm{d}\epsilon \\ &= \frac{1}{1 - 2Q(\eta_i)} \int_{-\eta_i}^{\eta_i} \frac{\epsilon^2}{\sqrt{2\pi}} \exp(-\epsilon^2/2)\, \mathrm{d}\epsilon \\ &= 1 - \sqrt{\frac{2}{\pi}} \frac{\eta_i \exp\left(-\eta_i^2/2\right)}{1 - 2Q(\eta_i)}.\end{aligned} \tag{13}$$

Therefore, from (12) and (13) and $\left(\sigma^i_k\right)^2 = \mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\mathbf{c}_i + r_i$, $\mathbf{k}^i_k = \mathbf{P}^{i-1}_{k|k}\mathbf{c}_i / \left(\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\mathbf{c}_i + r_i\right)$, one arrives at

$$\begin{aligned}\mathbf{P}^i_{k|k} &= \left[\left(\mathbf{I}_n - \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\right)\mathbf{P}^{i-1}_{k|k}\left(\mathbf{I}_n - \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\right)^{\mathrm{T}} + \mathbf{k}^i_k r_i (\mathbf{k}^i_k)^{\mathrm{T}}\right] \\ &\quad + \left(\sigma^i_k\right)^2\mathbf{k}^i_k\left[1 - \sqrt{\frac{2}{\pi}} \frac{\eta_i \exp\left(-\eta_i^2/2\right)}{1 - 2Q(\eta_i)}\right]\left(\mathbf{k}^i_k\right)^{\mathrm{T}} \\ &= \left[\mathbf{P}^{i-1}_{k|k} - \mathbf{P}^{i-1}_{k|k}\mathbf{c}_i\left(\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\mathbf{c}_i + r_i\right)^{-1}\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\right] \\ &\quad - \left(1 - \sqrt{\frac{2}{\pi}} \frac{\eta_i \exp\left(-\eta_i^2/2\right)}{1 - 2Q(\eta_i)}\right) \\ &\quad \times \mathbf{P}^{i-1}_{k|k}\mathbf{c}_i\left(\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\mathbf{c}_i + r_i\right)^{-1}\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k} \\ &= \mathbf{P}^{i-1}_{k|k} - \sqrt{\frac{2}{\pi}} \frac{\eta_i \exp\left(-\eta_i^2/2\right)}{1 - 2Q(\eta_i)} \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}.\end{aligned}$$

To write the discussed two scenarios in a more compact form, we arrive at

$$\begin{aligned}\mathbf{P}^i_{k|k} &= \left[\gamma^i_k + \left(1 - \gamma^i_k\right)\beta^i_k\right]\left(\mathbf{P}^{i-1}_{k|k} - \mathbf{k}^i_k\mathbf{c}_i\mathbf{P}^{i-1}_{k|k}\right) \\ &\quad + \left(1 - \gamma^i_k\right)\left(1 - \beta^i_k\right)\left(\mathbf{P}^{i-1}_{k|k} - \sqrt{\frac{2}{\pi}} \frac{\eta_i \exp\left(-\eta_i^2/2\right)}{1 - 2Q(\eta_i)} \mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\right) \\ &= \left[\gamma^i_k + \left(1 - \gamma^i_k\right)\beta^i_k + \left(1 - \gamma^i_k\right)\left(1 - \beta^i_k\right)\right]\mathbf{P}^{i-1}_{k|k} \\ &\quad - \Big[\gamma^i_k + \left(1 - \gamma^i_k\right)\beta^i_k - \left(1 - \gamma^i_k\right)\left(1 - \beta^i_k\right) \\ &\quad \times \sqrt{\frac{2}{\pi}} \frac{\eta_i \exp\left(-\eta_i^2/2\right)}{1 - 2Q(\eta_i)}\Big]\mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k} \\ &= \mathbf{P}^{i-1}_{k|k} - t\left(\gamma^i_k, \beta^i_k\right)\mathbf{k}^i_k\mathbf{c}_i^{\mathrm{T}}\mathbf{P}^{i-1}_{k|k}\end{aligned} \tag{14}$$

which completes the proof. □

Under the assumption of Algorithm 2, the sequences $\{\gamma^1_k\}_0^\infty$, $\{\gamma^2_k\}_0^\infty, \ldots, \{\gamma^m_k\}_0^\infty$ are independent and identically distributed (i.i.d.) processes with the approximately Gaussian pdf [15]. Further, assume the sequences $\{\beta^1_k\}_0^\infty$, $\{\beta^2_k\}_0^\infty, \ldots, \{\beta^m_k\}_0^\infty$ are mutually independent Bernoulli i.i.d. processes. Define $N_k = \mathrm{diag}\left[t(\gamma^1_k), t(\gamma^2_k), \ldots, t(\gamma^m_k)\right]$. For $i = 1, 2, \ldots, m$, let

$$\mu_i = \mathbb{E}[\gamma^i_k] = 2Q(\eta_i) \tag{15}$$

$$\nu_i = \frac{\eta_i \exp(\eta_i^2/2)}{\sqrt{\pi/2}(1 - 2Q(\eta_i))} \tag{16}$$

and $\mathbb{E}[\beta^i_k] = \beta$; in addition,

$$\begin{aligned}\mathbb{E}\left[l(\beta^i_k)\right] &= \mathbb{E}\left[\beta^i_k + (1 - \beta^i_k) \times \frac{\eta_i \exp(-\eta_i^2/2)}{\sqrt{\pi/2}(1 - 2Q(\eta_i))}\right] \\ &= \beta + (1 - \beta)\nu_i = \xi_i,\end{aligned} \tag{17}$$

$$\mathbb{E}\left[t(\gamma^i_k, \beta^i_k)\right] = \mathbb{E}\left[\gamma^i_k + (1 - \gamma^i_k)l(\beta^i_k)\right] = \mu_i + (1 - \mu_i)\xi_i = \lambda_i \tag{18}$$

where, in fact, we have $\nu_i = 1 - (1/\sqrt{2\pi})\int_{-\eta_i}^{\eta_i} \exp(-t^2/2)\,\mathrm{d}t \in [0, 1]$ and $\xi_i \in [0, 1]$. Moreover, ν_i is one strictly decreasing function in threshold η_i; this makes sense since the greater the threshold is, the less information will be transmitted with high power. Then, one can easily verify that $0 \le \lambda_i \le 1$, and therefore, λ_i can be somewhat physically interpreted as the normalised average information received by the FC resulting from the power scheduling and networked effect on transmitting y_k^i (and $1 - \lambda_i$ quantifies the corresponding average information loss rate). All λ_is together will govern the mean-square stability of estimation error covariance matrix, which will be investigated in the ensuing section. Therefore, we will refer to λ_i hereafter other than the specific parameters η_i, β.

4 Statistical properties of error covariance matrix iterations

In this section, we will focus on the statistical properties of the error covariance matrix. Denote $\{\gamma_k^i\} := \{\{\gamma_k^1\}_0^\infty, \{\gamma_k^2\}_0^\infty, \ldots, \{\gamma_k^m\}_0^\infty\}$ and $\{\beta_k^i\} := \{\{\beta_k^1\}_0^\infty, \{\beta_k^2\}_0^\infty, \ldots, \{\beta_k^m\}_0^\infty\}$. Before delving into main results, some preliminaries will be given in the following.

Let $\mathbb{S}_+^n = \{\mathbf{S} \in \mathbb{R}^{n\times n} | \mathbf{S} \ge \mathbf{0}\}$. Define the function $\mathbf{h}: \mathbb{S}_+^n \to \mathbb{S}_+^n$ and the function $\mathbf{g}_{\lambda_i}: \mathbb{S}_+^n \to \mathbb{S}_+^n$ as follows:

$$\mathbf{h}(\mathbf{X}) \triangleq \mathbf{A}\mathbf{X}\mathbf{A}^{\mathrm{T}} + \mathbf{Q} \tag{19}$$

$$\mathbf{g}_{\lambda_i}(\mathbf{X}) \triangleq \mathbf{X} - \lambda_i \mathbf{X}\mathbf{c}_i\left(\mathbf{c}_i^{\mathrm{T}}\mathbf{X}\mathbf{c}_i + r_i\right)^{-1}\mathbf{c}_i^{\mathrm{T}}\mathbf{X} \tag{20}$$

$$\mathbf{g}_{\lambda_i} \circ \boldsymbol{h}(\mathbf{X}) \triangleq \mathbf{g}_{\lambda_i}(\mathbf{h}(\mathbf{X})) \tag{21}$$

and here the notation $\circ$ denotes the function composition. Hereafter, the composition notation will be ignored as $\mathbf{g}_{\lambda_i} \circ \boldsymbol{h}(\mathbf{X}) = \mathbf{g}_{\lambda_i}\boldsymbol{h}(\mathbf{X})$ if no confusion raises. Therefore, the covariance update in the proposed KF formulation in Algorithm 2 becomes

$$\begin{aligned}
\mathbf{P}_{k|k-1} &= \mathbf{h}(\mathbf{P}_{k-1|k-1}) \\
\mathbf{P}_{k|k}^1 &= \mathbf{g}_{\lambda_1}(\mathbf{P}_{k|k}^0) = \mathbf{g}_{\lambda_1}(\mathbf{P}_{k|k-1}) \\
\mathbf{P}_{k|k}^i &= \mathbf{g}_{\lambda_i}(\mathbf{P}_{k|k}^{i-1}), \quad 2 \le i \le m-1, \\
\mathbf{P}_{k|k} = \mathbf{P}_{k|k}^m &= \mathbf{g}_{\lambda_m}(\mathbf{P}_{k|k}^{m-1}).
\end{aligned}$$

Define also

$$\mathbf{P}_{k|k-1} = \mathbf{h}(\mathbf{P}_{k-1|k-1}) \tag{22}$$

$$\mathbf{P}_{k|k} = \mathscr{M}_m(\mathbf{P}_{k|k-1}) \triangleq \mathbf{g}_{\lambda_m}\mathbf{g}_{\lambda_{m-1}}\cdots\mathbf{g}_{\lambda_1}(\mathbf{P}_{k|k-1}). \tag{23}$$

Denote the function $\boldsymbol{\varphi}: \mathbb{S}^n \to \mathbb{S}^n$ by the transformation from $\mathbf{P}_{k-1|k-1}$ to $\mathbf{P}_{k|k}$, namely,

$$\mathbf{P}_{k|k} = \boldsymbol{\varphi}(\mathbf{P}_{k-1|k-1}) \triangleq \mathbf{M}_m h(\mathbf{P}_{k-1|k-1}). \tag{24}$$

To analyse the convergence of the estimation error covariance matrix, we define the modified algebraic Riccati equation (MARE) in the following way

$$\boldsymbol{\varphi}(\mathbf{P}_k) = \mathbf{g}_{\lambda_m}\mathbf{g}_{\lambda_{m-1}}\cdots\mathbf{g}_{\lambda_1}\mathbf{h}(\mathbf{P}_k) \tag{25}$$

where we used the simplified notation $\mathbf{P}_k = \mathbf{P}_{k|k}, k \ge 0$. Meanwhile, as explained, the covariance matrices $\{\mathbf{P}_k\}_0^\infty$ depend non-linearly on the specific realisation of the stochastic processes $\{\gamma_k^i\}$ and $\{\beta_k^i\}$, hence the proposed KF is stochastic and cannot be determined offline. As a result, only statistical properties with respect to the covariance matrices of the proposed KF can therefore be established.

Before we will formally study convergence properties of the MARE in (25), let us introduce some supporting lemmas. The first one delineates some basic properties of an auxiliary function, which provide prerequisites of proving stability of MARE [24].

Lemma 1 [24]: Let the function $\boldsymbol{\psi}_{\lambda_i}$ for $i = 1, \ldots, m$ be

$$\boldsymbol{\psi}_{\lambda_i}(\mathbf{l}_i, \mathbf{X}) = (1 - \lambda_i)\mathbf{X} + \lambda_i\left(\mathbf{E}_i\mathbf{X}\mathbf{E}_i^{\mathrm{T}} + \mathbf{l}_i r_i \mathbf{l}_i^{\mathrm{T}}\right) \tag{26}$$

where $\mathbf{E}_i = \mathbf{I}_n + \mathbf{l}_i\mathbf{c}_i^{\mathrm{T}}, r_i > 0, \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{S}_+^n$. Then the following facts hold:

1. With given $\mathbf{l}_i^{\mathbf{X}} = -\mathbf{X}\mathbf{c}_i(\mathbf{c}_i^{\mathrm{T}}\mathbf{X}\mathbf{c}_i + r_i)^{-1}$, $\mathbf{g}_{\lambda_i}(\mathbf{X}) = \boldsymbol{\psi}_{\lambda_i}(\mathbf{l}_i^{\mathbf{X}}, \mathbf{X})$
2. $\mathbf{g}_{\lambda_i}(\mathbf{X}) = \min_{\mathbf{l}_i} \boldsymbol{\psi}(\mathbf{l}_i, \mathbf{X}) \le \boldsymbol{\psi}(\mathbf{l}_i, \mathbf{X}), \forall \mathbf{l}_i$
3. If $\mathbf{X} \le \mathbf{Y}$, then $\mathbf{g}_{\lambda_i}(\mathbf{X}) \le \mathbf{g}_{\lambda_i}(\mathbf{Y})$
4. If $\lambda_i \ge \lambda_j$, then $\mathbf{g}_{\lambda_i}(\mathbf{X}) \le \mathbf{g}_{\lambda_j}(\mathbf{X})$
5. If $\tau \in [0, 1]$, then $\mathbf{g}_{\lambda_i}(\tau\mathbf{X} + (1 - \tau)\mathbf{Y}) \ge \tau\mathbf{g}_{\lambda_i}(\mathbf{X}) + (1 - \tau)\mathbf{g}_{\lambda_i}(\mathbf{Y})$.

Proof: The proofs are analogous to those of Lemma 1 in [24] with appropriate notation adaptations. □

Note that the relationship between the functions $\mathbf{g}_{\lambda_i}$ and $\boldsymbol{\psi}_{\lambda_i}$ has been established. In order to investigate the convergence properties of the MARE in (25), the relationship between the composite function $\mathbf{g}_{\lambda_m}\mathbf{g}_{\lambda_{m-1}}\cdots\mathbf{g}_{\lambda_1}$ and the auxiliary function $\boldsymbol{\psi}_{\lambda_m}\boldsymbol{\psi}_{\lambda_{m-1}}\cdots\boldsymbol{\psi}_{\lambda_1}$ will be detailed next.

Let us introduce the function $\mathscr{T}_s(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_s, \mathbf{X}) = \boldsymbol{\psi}_{\lambda_s}\boldsymbol{\psi}_{\lambda_{s-1}}\cdots\boldsymbol{\psi}_{\lambda_1}(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_s, \mathbf{X})$, $1 \le s \le m$. Then, (see (27)) where, to make the expression more concrete, we defined $\lambda_0 = 1$ and $\mathbf{E}_0 = \mathbf{I}_n, r_0 = 0, \mathscr{T}_{-1} = \mathbf{X}, \mathscr{T}_0(\mathbf{X}) = \mathbf{X}$.

For ease of exposition, denote

$$\begin{aligned}
\eta_{j,s}^2 &= \prod_{i=j+1}^{s}(1 - \lambda_i)\lambda_j, \quad 0 \le j \le s-1 \\
\eta_{s,s}^2 &= \lambda_s.
\end{aligned} \tag{28}$$

More importantly, it is easy to exploit the fact that the sum of $s + 1$ coefficients $\eta_{j,s}^2$, $0 \le j \le s$, is identically 1, i.e.

$$\sum_{j=0}^{s}\eta_{j,s}^2 = \sum_{j=0}^{s-1}\Big[\prod_{i=j+1}^{s}(1 - \lambda_i)\lambda_j\Big] + \eta_{s,s}^2 = (1 - \lambda_s) + \lambda_s = 1.$$

$$\begin{aligned}
\mathscr{T}_s &= \boldsymbol{\psi}_{\lambda_s}(\mathbf{l}_s, \mathscr{T}_{s-1}) \\
&= (1 - \lambda_s)\mathscr{T}_{s-1} + \lambda_s\left(\mathbf{E}_s\mathscr{T}_{s-1}\mathbf{E}_s^{\mathrm{T}} + \mathbf{l}_s r_s \mathbf{l}_s^{\mathrm{T}}\right) \\
&= \sum_{j=1}^{s-1}\prod_{i=j+1}^{s}(1 - \lambda_i)\lambda_j\left(\mathbf{E}_j\mathscr{T}_{j-1}\mathbf{E}_j^{\mathrm{T}} + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}}\right) + \prod_{i=1}^{s}(1 - \lambda_i)\mathbf{X} + \lambda_s\left(\mathbf{E}_s\mathscr{T}_{s-1}\mathbf{E}_s^{\mathrm{T}} + \mathbf{l}_s r_s \mathbf{l}_s^{\mathrm{T}}\right) \\
&= \sum_{j=0}^{s-1}\prod_{i=j+1}^{s}(1 - \lambda_i)\lambda_j\left(\mathbf{E}_j\mathscr{T}_{j-1}\mathbf{E}_j^{\mathrm{T}} + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}}\right) + \lambda_s\left(\mathbf{E}_s\mathscr{T}_{s-1}\mathbf{E}_s^{\mathrm{T}} + \mathbf{l}_s r_s \mathbf{l}_s^{\mathrm{T}}\right)
\end{aligned} \tag{27}$$

Alternatively, (27) can be given by

$$\begin{aligned}&\mathscr{T}_{-1} = \mathbf{X}, \mathscr{T}_0 = \mathbf{X},\\ &\mathscr{T}_s = \sum_{j=0}^{s} \eta_{j,s}^2\left(\mathbf{E}_j\mathscr{T}_{j-1}\mathbf{E}_j^{\mathrm{T}} + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}}\right), \quad 1 \le s \le m.\end{aligned} \tag{29}$$

Therefore,

$$\mathscr{T}_m = \sum_{j=0}^{m} \eta_{j,m}^2\left(\mathbf{E}_j\mathscr{T}_{j-1}\mathbf{E}_j^{\mathrm{T}} + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}}\right) \tag{30}$$

where $\eta_{0,0}^2 = 1, \mathbf{E}_0 = \mathbf{I}_n, r_0 = 0$ and $\mathbf{X} \ge \mathbf{0}$, and $\mathscr{T}_j$ is defined in (29) with $\eta_{j,m}$ given by (28).

Likewise in Lemma 1, we study basic properties of the function $\mathbf{T}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X})$, which will be presented in form of lemmas in the following.

Lemma 2: Consider the function $\mathbf{T}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X})$ as stated by (30) with $\mathbf{E}_j = \mathbf{I}_n + \mathbf{l}_j\mathbf{c}_j^{\mathrm{T}}$. Assume $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{S}_+^n$. Then, the following facts hold:

1. With given $\mathbf{l}_j^X = -\mathscr{T}_{j-1}^{\mathbf{X}}\mathbf{c}_j\left(\mathbf{c}_j^{\mathrm{T}}\mathbf{T}_{j-1}^{\mathbf{X}}\mathbf{c}_j + r_j\right)^{-1}, j = 1, 2, \ldots, m, \mathbf{M}_m(\mathbf{X}) = \mathbf{T}_m, \left(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_m^{\mathbf{X}}, \mathbf{X}\right)$ where $\mathscr{T}_{j-1}^{\mathbf{X}} = \mathscr{T}_{j-1}(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_{j-1}^{\mathbf{X}}, \mathbf{X})$
2. $\mathbf{M}_m(\mathbf{X}) = \min_{\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m} \mathscr{T}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X}) \le \mathscr{T}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X}), \forall \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m \in \mathbb{R}^{n\times 1}$
3. If $\mathbf{X} \le \mathbf{Y}$, then $\mathbf{M}_m(\mathbf{X}) \le \mathbf{M}_m(\mathbf{Y})$
4. If $\tau \in [0, 1]$, then $\mathbf{M}_m(\tau\mathbf{X} + (1-\tau)\mathbf{Y}) \ge \tau\mathbf{M}_m(\mathbf{X}) + (1-\tau)\mathbf{M}_m(\mathbf{Y})$
5. $\mathscr{M}_m(\mathbf{X}) \ge \prod_{j=1}^m (1-\lambda_j)\mathbf{X}$
6. For a random variable $\mathbf{X}$, $\prod_{j=1}^m (1-\lambda_j)\mathbb{E}[\mathbf{X}] \le \mathbb{E}[\mathbf{M}_m(\mathbf{X})] \le \mathscr{M}_m(\mathbb{E}[\mathbf{X}])$.

To account for $\mathbf{P}_{k+1|k} = h(\mathbf{P}_{k|k})$, the auxiliary function $\boldsymbol{\phi}_m$ can be given in the following way:

$$\begin{aligned}&\boldsymbol{\phi}_{-1} = h(\mathbf{X}), \boldsymbol{\phi}_0 = h(\mathbf{X}),\\ &\boldsymbol{\phi}_s = \sum_{j=0}^{s} \eta_{j,s}^2\left(\mathbf{E}_j\boldsymbol{\phi}_{j-1}\mathbf{E}_j^{\mathrm{T}} + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}}\right), \quad s = 1, 2, \ldots, m-1,\\ &\boldsymbol{\phi}_m = \sum_{j=0}^{m} \eta_{j,m}^2\left(\mathbf{E}_j\boldsymbol{\phi}_{j-1}\mathbf{E}_j^{\mathrm{T}} + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}}\right).\end{aligned} \tag{31}$$

In the ensuing part, we will present some useful properties of $\boldsymbol{\phi}_m(\mathbf{X})$, where $\mathbf{X} \ge \mathbf{0}$. These properties allow us to find a lower and upper bound for the steady state error covariance matrix $\lim_{k\to\infty} \mathbb{E}[\mathbf{P}_k]$ which is independent of the initial condition $\mathbf{P}_0$.

Lemma 3: Consider the function $\boldsymbol{\phi}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, X)$ as stated by (31) with $\mathbf{E}_j = \mathbf{I}_n + \mathbf{l}_j\mathbf{c}_j^{\mathrm{T}}$. Assume $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbb{S}_+^n$. Then, the following facts hold:

1. With given $\mathbf{l}_j^{\mathbf{X}} = -\boldsymbol{\phi}_{j-1}^{\mathbf{X}}\mathbf{c}_j\left(\mathbf{c}_j^{\mathrm{T}}\boldsymbol{\phi}_{j-1}^{\mathbf{X}}\mathbf{c}_j + r_j\right)^{-1}$, $j = 1, 2, \ldots, m, \boldsymbol{\varphi}(X) = \boldsymbol{\phi}_m\left(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_m^{\mathbf{X}}, \mathbf{X}\right)$, where $\boldsymbol{\phi}_{j-1}^{\mathbf{X}} = \boldsymbol{\phi}_{j-1}(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_{j-1}^{\mathbf{X}}, \mathbf{X})$
2. $\boldsymbol{\varphi}(X) = \min_{\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m} \boldsymbol{\phi}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X}) \le \boldsymbol{\phi}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X})$, $\forall \mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m \in \mathbb{R}^{n\times 1}$
3. If $\mathbf{X} \le \mathbf{Y}$, then $\boldsymbol{\varphi}(\mathbf{X}) \le \boldsymbol{\varphi}(\mathbf{Y})$
4. If $\tau \in [0, 1]$, then $\boldsymbol{\varphi}(\tau\mathbf{X} + (1-\tau)\mathbf{Y}) \ge \tau\boldsymbol{\varphi}(\mathbf{X}) + (1-\tau)\boldsymbol{\varphi}(\mathbf{Y})$
5. $\boldsymbol{\varphi}(\mathbf{X}) \ge \prod_{j=1}^m (1-\lambda_j)\mathbf{A}\mathbf{X}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$
6. If $\bar{\mathbf{X}} \ge \boldsymbol{\varphi}(\bar{\mathbf{X}})$, then $\bar{\mathbf{X}} > \mathbf{0}$
7. For a random variable $\mathbf{X}$, $\prod_{j=1}^m (1-\lambda_j)\mathbf{A}\mathbb{E}[\mathbf{X}]\mathbf{A}^{\mathrm{T}} + \mathbf{Q} \le \mathbb{E}[\boldsymbol{\varphi}(\mathbf{X})] \le \boldsymbol{\varphi}(\mathbb{E}[\mathbf{X}])$.

Remark 3: Observe that if we substitute $\mathbf{X} = \mathbf{P}_{k|k}$ into Fact (7) in Lemma 3, it follows that $\prod_{j=1}^m (1-\lambda_j)\mathbf{A}\mathbb{E}[\mathbf{P}_k]\mathbf{A}^{\mathrm{T}} + \mathbf{Q} \le \mathbb{E}[\boldsymbol{\varphi}(\mathbf{P}_k)] \le \boldsymbol{\varphi}(\mathbb{E}[\mathbf{P}_k])$. Since $\mathbb{E}[\mathbf{P}_{k+1}|\mathbf{P}_k] = \boldsymbol{\varphi}(\mathbf{P}_k)$ and $\mathbb{E}[\mathbf{P}_{k+1}] = \mathbb{E}[\boldsymbol{\varphi}(\mathbf{P}_k)]$, then $\prod_{j=1}^m (1-\lambda_j)A\mathbb{E}[\mathbf{P}_k]A' + Q \le \mathbb{E}[\mathbf{P}_{k+1}] \le \boldsymbol{\varphi}(\mathbb{E}[\mathbf{P}_k])$. That is, the expected value of $\mathbf{P}_{k+1|k}$ can be lower-bounded and upper-bounded by $\prod_{j=1}^m (1-\lambda_j)\mathbf{A}\mathbb{E}[\mathbf{P}_k]\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$ and $\boldsymbol{\varphi}(\mathbb{E}[\mathbf{P}_k])$ both as functions of $\mathbb{E}[\mathbf{P}_k]$, respectively.

To facilitate the convergence analysis, let us define the linear part of function $\boldsymbol{\phi}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X})$ in terms of variable $\mathbf{X}$ as another auxiliary function, namely

$$\mathscr{L}_m(\mathbf{Y}) = \sum_{j=0}^{m} \eta_{j,m}^2\left(\mathbf{E}_j\boldsymbol{\phi}_{j-1}\mathbf{E}_j^{\mathrm{T}}\right) \tag{32}$$

where $\boldsymbol{\phi}_{j-1}, j = 0, 1, \ldots, m$ are defined in (31). Then, the following lemma can be readily presented.

Lemma 4: Consider the function $\mathbf{L}_m(\mathbf{Y})$ as stated in (32). If there exists a positive definite matrix $\bar{\mathbf{Y}} > \mathbf{0}$ such that $\bar{\mathbf{Y}} > \mathbf{L}_m(\bar{\mathbf{Y}})$, then

1. $\forall \mathbf{W} \ge \mathbf{0}$, $\lim_{k\to\infty} \mathbf{L}_m^k(\mathbf{W}) = \mathbf{0}$
2. Given $\mathbf{U} > \mathbf{0}$, let the following sequence

$$\mathbf{Y}_{k+1} = \mathbf{L}_m(\mathbf{Y}_k) + \mathbf{U}$$

initialised at $\mathbf{Y}_0 \ge \mathbf{0}$. Then, the sequence $\mathbf{Y}_k$ is bounded.

Lemma 5: Consider the function $\boldsymbol{\phi}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{X})$ defined in (31). Assume there exist m gain matrices $\bar{\mathbf{l}}_1, \bar{\mathbf{l}}_2, \ldots, \bar{\mathbf{l}}_m$ and a positive definite matrix $\bar{\mathbf{P}}$ such that

$$\bar{\mathbf{P}} > \mathbf{0} \text{ and } \bar{\mathbf{P}} > \boldsymbol{\phi}_m(\bar{\mathbf{l}}_1, \bar{\mathbf{l}}_2, \ldots, \bar{\mathbf{l}}_m, \bar{\mathbf{P}}).$$

Then, the sequence $\mathbf{P}_k = \boldsymbol{\varphi}^k(\mathbf{P}_0)$ is bounded for any given $\mathbf{P}_0$. That is, there exists a positive definite matrix $\mathbf{M}_{\mathbf{P}_0} > \mathbf{0}$ depending on $\mathbf{P}_0$ such that

$$\mathbf{P}_k \le \mathbf{M}_{\mathbf{P}_0}, \quad \forall k \ge 0.$$

Lemma 6: Let $\mathbf{Y}_{s+1} = f(\mathbf{Y}_s)$ and $\mathbf{Z}_{s+1} = f(\mathbf{Z}_s)$. Suppose that the function $f(\mathbf{Y})$ is monotonically increase in $\mathbf{Y}$. Then:

$$\begin{aligned}&\mathbf{Y}_1 \ge \mathbf{Y}_0 \Longrightarrow \mathbf{Y}_{s+1} \ge \mathbf{Y}_s, \quad \forall s \ge 0\\ &\mathbf{Y}_1 \le \mathbf{Y}_0 \Longrightarrow \mathbf{Y}_{s+1} \le \mathbf{Y}_s, \quad \forall s \ge 0\\ &\mathbf{Y}_0 \le \mathbf{Z}_0 \Longrightarrow \mathbf{Y}_k \le \mathbf{Z}_s, \quad \forall s \ge 0.\end{aligned}$$

5 Sufficient and necessary convergence conditions

After establishing these lemmas, we are ready to present our sufficient and necessary conditions for the mean-squared stability of the averaged estimation error covariance matrix.

Theorem 1 (Sufficient condition): Consider the function $\boldsymbol{\phi}_m = \sum_{j=0}^m \eta_{j,m}^2(\mathbf{E}_j\boldsymbol{\phi}_{j-1}\mathbf{E}_j^{\mathrm{T}} + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}})$ defined in (31), where $\{\eta_{j,m}\}$

are defined in (28). If there exist m matrices $\tilde{\mathbf{l}}_j, j = 1, 2, \ldots, m$ and a positive definite matrix $\bar{\mathbf{P}}$ such that

$$\bar{\mathbf{P}} > \mathbf{0} \text{ and } \bar{\mathbf{P}} > \boldsymbol{\phi}_m(\tilde{\mathbf{L}}_1, \tilde{\mathbf{L}}_2, \ldots, \tilde{\mathbf{L}}_m, \bar{\mathbf{P}}). \tag{33}$$

Then, the following facts are true:

1. The MARE converges for any initial condition $\mathbf{P}_0 \geq \mathbf{0}$ and the limit

$$\lim_{t \to \infty} \mathbf{P}_k = \lim_{k \to \infty} \boldsymbol{\phi}_m^k(\mathbf{P}_0) = \bar{\mathbf{P}}$$

is independent of the initial condition $\mathbf{P}_0$.

2. $\bar{\mathbf{P}}$ is the unique positive definite fixed point of the MARE.

Proof:

i. To begin with, we verify the convergence of the MARE sequence initialised at $\mathbf{Q}_0 = \mathbf{0}$ and therefore $\mathbf{Q}_k = \boldsymbol{\varphi}^k(\mathbf{0})$. Then it directly follows that $\mathbf{0} = \mathbf{Q}_0 \leq \boldsymbol{\varphi}(\mathbf{0}) = \mathbf{Q}_1$, and in the light of Fact (3) in Lemma 3, it gives that

$$\mathbf{Q}_1 = \boldsymbol{\varphi}(\mathbf{Q}_0) \leq \boldsymbol{\varphi}(\mathbf{Q}_1) = \mathbf{Q}_2.$$

From Lemmas 5 and 6, a monotonically non-decreasing sequence of matrices follow directly from a simple inductive argument and the sequence is also upper-bounded, that is,

$$\mathbf{0} = \mathbf{Q}_0 \leq \mathbf{Q}_1 \leq \mathbf{Q}_2 \leq \cdots \leq \mathbf{M}_{\mathbf{Q}_0}.$$

Here, one can easily verify that the monotonically non-decreasing and upper-bounded sequence converges from the Bolzano–Weierstrass theorem, that is, $\lim_{k \to \infty} \mathbf{Q}_k = \bar{\mathbf{P}}$ where $\bar{\mathbf{P}} \geq \mathbf{0}$ is a fixed point of the following modified Riccati iteration

$$\bar{\mathbf{P}} = \boldsymbol{\varphi}(\bar{\mathbf{P}}). \tag{34}$$

Then, we show that the modified Riccati iteration initialised at $\mathbf{S}_0 \geq \bar{\mathbf{P}}$ also converges to the same point $\bar{\mathbf{P}}$. By resorting to (32), one obtains

$$\bar{\mathbf{P}} = \boldsymbol{\varphi}(\bar{\mathbf{P}}) = \mathbf{L}_m^{\bar{\mathbf{P}}}(\bar{\mathbf{P}}) + \mathbf{Q} + \mathbf{N}_m^{\bar{\mathbf{P}}} > \mathbf{L}_m^{\bar{\mathbf{P}}}(\bar{\mathbf{P}})$$

where $\mathbf{L}_m^{\bar{\mathbf{P}}}(\mathbf{Y}) = \sum_{j=0}^{m} \eta_{j,m}^2 \mathbf{E}_j^{\bar{\mathbf{P}}} \boldsymbol{\phi}_{j-1}(\mathbf{E}_j^{\bar{\mathbf{P}}})^{\mathrm{T}}$. Therefore, $\mathbf{L}_m^{\bar{\mathbf{P}}}$ satisfies the condition of Lemma 4. Accordingly, we realise that

$$\lim_{k \to \infty} (\mathbf{L}_m^{\bar{\mathbf{P}}})^k(\mathbf{Y}) = \mathbf{0}, \quad \forall \mathbf{Y} \geq \mathbf{0}.$$

Assume that $\mathbf{S}_0 \geq \bar{\mathbf{P}}$ and then,

$$\mathbf{S}_1 = \boldsymbol{\varphi}(\mathbf{S}_0) \geq \boldsymbol{\varphi}(\bar{\mathbf{P}}) = \bar{\mathbf{P}}$$

where is due to the monotonic increase property of $\boldsymbol{\varphi}(\mathbf{X})$ and (34). By induction,

$$\mathbf{S}_k \geq \bar{\mathbf{P}}, \quad \forall k > 0.$$

Meanwhile, we have

$$\begin{aligned}
\mathbf{0} \leq \mathbf{S}_{k+1} - \bar{\mathbf{P}} &= \boldsymbol{\varphi}(\mathbf{S}_k) - \boldsymbol{\varphi}(\bar{\mathbf{P}}) \\
&= \boldsymbol{\phi}_m(\mathbf{l}_1^{\mathbf{S}_k}, \mathbf{l}_2^{\mathbf{S}_k}, \ldots, \mathbf{l}_m^{\mathbf{S}_k}, \mathbf{S}_k) - \boldsymbol{\phi}_m(\mathbf{l}_1^{\bar{P}}, \mathbf{l}_2^{\bar{\mathbf{P}}}, \ldots, \mathbf{l}_m^{\bar{\mathbf{P}}}, \bar{\mathbf{P}}) \\
&\leq \boldsymbol{\phi}_m(\mathbf{l}_1^{\bar{\mathbf{P}}}, \mathbf{l}_2^{\bar{\mathbf{P}}}, \ldots, \mathbf{l}_m^{\bar{\mathbf{P}}}, \mathbf{S}_k) - \boldsymbol{\phi}_m(\mathbf{l}_1^{\bar{\mathbf{P}}}, \mathbf{l}_2^{\bar{\mathbf{P}}}, \ldots, \mathbf{l}_m^{\bar{\mathbf{P}}}, \bar{\mathbf{P}}) \\
&= \sum_{j=0}^{m} \eta_{j,m}^2 \mathbf{E}_j^{\bar{\mathbf{P}}} \left(\boldsymbol{\phi}_j^{\mathbf{S}_k} - \boldsymbol{\phi}_j^{\bar{\mathbf{P}}}\right) (\mathbf{E}_j^{\bar{\mathbf{P}}})^{\mathrm{T}} \\
&= \mathbf{L}_m^{\bar{\mathbf{P}}}(\mathbf{S}_k - \bar{\mathbf{P}}).
\end{aligned}$$

Then, since $\lim_{k \to \infty} \mathscr{L}_m^{\bar{\mathbf{P}}}(\mathbf{S}_k - \bar{\mathbf{P}}) = \mathbf{0}$, it directly follows that $\lim_{k \to \infty} (\mathbf{S}_{k+1} - \bar{\mathbf{P}}) = \mathbf{0}$. That is, we have shown $\mathbf{S}_k \to \bar{\mathbf{P}}$ as $k \to \infty$ when $\mathbf{S}_0 \geq \bar{\mathbf{P}}$.

In the following, we are ready to justify that the modified Riccati iteration $\mathbf{P}_k = \boldsymbol{\varphi}^k(\mathbf{P}_0)$ converges to $\bar{\mathbf{P}}$ for all initial conditions $\mathbf{P}_0 \geq \mathbf{0}$. Let $\mathbf{Q}_0 = \mathbf{0}$ and also $\mathbf{S}_0 = \bar{\mathbf{P}} + \mathbf{P}_0$. Then consider the three Riccati iterations initialised at $\mathbf{Q}_0$, $\mathbf{P}_0$ and $\mathbf{S}_0$, respectively. Clearly, $\mathbf{Q}_0 \leq \mathbf{P}_0 \leq \mathbf{S}_0$, and appealing to Lemma 6, it gives that $\mathbf{0} \leq \mathbf{Q}_k \leq \mathbf{P}_k \leq \mathbf{S}_k$, $\forall k \geq 0$. Given that both the sequence $\mathbf{Q}_k$ and the sequence $\mathbf{S}_k$ converge to $\bar{\mathbf{P}}$, consequently, we have $\lim_{k \to \infty} \mathbf{P}_k = \bar{\mathbf{P}}$.

ii. Let us further postulate there exists another positive semi-definite matrix $\hat{\mathbf{P}} \geq \mathbf{0}$ such that $\hat{P} = \boldsymbol{\varphi}(\hat{\mathbf{P}})$. Let us consider the Riccati iteration initialised at $\hat{\mathbf{P}}$, and therefore, we can derive the following sequence

$$\hat{\mathbf{P}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}, \ldots.$$

Clearly, every Riccati iteration is shown to converge to the same limit $\bar{\mathbf{P}}$. Therefore, we have $\hat{\mathbf{P}} = \bar{\mathbf{P}}$.

□

In the sequel, a toy example of a scalar-state vector-observation system is provided to justify the existence of sufficient condition in Theorem 1.

Example: We consider the following scalar-state vector-observation system [31]

$$\begin{aligned}
x_{k+1} &= a x_k + \omega_k \\
\mathbf{y}_k &= \mathbf{c} x_k + \boldsymbol{\nu}_k
\end{aligned}$$

where $a = 1.2$, $\mathbf{c} = [c_1, c_2]' = [1, 1]'$, noise covariances are $q = 1$ and $\mathbf{R} = \mathrm{diag}\{r_1, r_2\} = \mathrm{diag}\{0.1, 1\}$. For simplicity, consider $\lambda_1 = \lambda_2 = 0.6$, and let l_1, l_2 be, for instance, such that $l_1 = -1$, $-2.8276 < l_2 < 0.8276$ or $l_2 = -1$, $-2.8276 < l_1 < 0.8276$. Then one can always find $p > 0$ such that l_1, l_2, p satisfy condition (33) in Theorem 1. That is, the expected estimation error covariance matrix will converge.

In the ensuing part, we will present one necessary condition for ensuring mean-square stability of expected estimation error covariance matrix which extends the result in [15] to general linear systems with data packet drops.

Theorem 2 (Necessary condition): Consider system (1) and Algorithm 2. Assume that $\mathbf{A}$ is unstable, that $(\mathbf{A}, \mathbf{Q}^{1/2})$ is controllable and that $(\mathbf{C}, \mathbf{A})$ is observable. If $\mathrm{E}[\mathbf{P}_k] \leq \mathbf{M}_{\mathbf{P}_0}, \forall k \geq 0$ holds for any initial condition $\mathbf{P}_0 \geq \mathbf{0}$, then $\lambda_1, \lambda_2, \ldots, \lambda_m$ defined in (18) should satisfy the following condition

$$\prod_{i=1}^{m} (1 - \lambda_i) \leq \frac{1}{\left(\max_i |\sigma_i(\mathbf{A})|\right)^2} \tag{35}$$

where $\sigma_i(\mathbf{A}), i = 1, 2, \ldots, n$, are all eigenvalues of square matrix $\mathbf{A}$, and $\mathbf{M}_{\mathbf{P}_0} > \mathbf{0}$ depends on the initial condition $\mathbf{P}_0 \geq \mathbf{0}$.

Proof: The proof follows straightforwardly from Fact (7) in Lemma 3. □

6 Concluding remarks

In this paper we introduced a measurement-innovation based power scheduler for wireless sensors in terms of optimally deciding whether to use a high or low transmission-power to communicate scalar observations to the FC where estimation is performed. The high transmission-power is used to transmit the well-defined 'informative' measurements and the opposite for 'non-informative' ones. Further, the high power transmission power is assumed to lead to reliable data transmission while the low transmission power may cause data packet drops. Under this new setup, the MMSE estimator was derived. Convergence analysis of the averaged estimation error covariance matrix were provided, while sufficient and necessary conditions guaranteeing its convergence were established for general linear stochastic systems.

7 Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grant nos. 61104097, 61321002, 61120106010, 61522303, and U1509215), the Program for Changjiang Scholars and Innovative Research Team in University (IRT1208), the ChangJiang Scholars Program, the Beijing Outstanding Ph.D. Program Mentor Grant (20131000704), the Program for New Century Excellent Talents in University (NCET-13-0045), and the Beijing Higher Education Young Elite Teacher Project.

8 References

[1] Akyildiz, I., Su, W., Sankarasubramaniam, Y., *et al.*: 'Wireless sensor networks: a survey', *Comput. Netw.*, 2002, **38**, (4), pp. 393–422

[2] Olfati-Saber, R.: 'Distributed Kalman filtering for sensor networks'. Proc. 48th IEEE Conf. Decision and Control, New Orleans, LA, USA, December 2007, pp. 44–49

[3] Ribeiro, A., Giannakis, G.B., Roumeliotis, S.: 'SOI-KF: distributed Kalman filtering with low-cost communications using the sign of innovations', *IEEE Trans. Signal Process.*, 2006, **54**, (12), pp. 4782–4795

[4] Msechu, E., Giannakis, G.B.: 'Sensor-centric data reduction for estimation with WSNs via censoring and quantization', *IEEE Trans. Signal Process.*, 2012, **60**, (1), pp. 400–414

[5] Liu, Y., Xu, B.: 'Filter designing with finite packet losses and its application for stochastic systems', *IET Control Theory Appl.*, 2011, **5**, (6), pp. 775–784

[6] Hu, J., Wang, Z., Gao, H., *et al.*: 'Extended Kalman filtering with stochastic nonlinearities and multiple missing measurements', *Automatica*, 2012, **48**, (9), pp. 2007–2015

[7] Hu, J., Wang, Z., Shen, B., *et al.*: 'Gain-constrained recursive filtering with stochastic nonlinearities and probabilistic sensor delays', *IEEE Trans. Signal Process.*, 2013, **61**, (5), pp. 1230–1238

[8] Xiao, J.J., Cui, S., Luo, Z.Q., *et al.*: 'Power scheduling of universal decentralized estimation in sensor networks', *IEEE Trans. Signal Process.*, 2006, **54**, (2), pp. 413–422

[9] Dong, H., Wang, Z., Lam, J., *et al.*: 'Fuzzy-model-based robust fault detection with stochastic mixed time delays and successive packet dropouts', *IEEE Trans. Syst. Man Cybern. B*, 2012, **42**, (2), pp. 365–376

[10] You, K., Xie, L., Sun, S., *et al.*: 'Quantized filtering of linear stochastic systems', *Trans. Inst. Meas. Control*, 2011, **33**, (6), pp. 683–698

[11] Savage, C., Scala, B.: 'Optimal scheduling of scalar Gauss–Markov systems with a terminal cost function', *IEEE Trans. Autom. Control*, 2009, **54**, (5), pp. 1100–1105

[12] Shi, L., Cheng, P., Chen, J.: 'Sensor data scheduling for optimal state estimation with communication energy constraint', *Automatica*, 2011, **47**, (8), pp. 1693–1698

[13] Shi, L., Xie, L.: 'Optimal sensor power scheduling for state estimation of Gauss–Markov systems over a packet-dropping network', *IEEE Trans. Signal Process.*, 2012, **60**, (5), pp. 2701–2705

[14] Suh, Y., Nguyen, V., Ro, Y.: 'Modified Kalman filter for networked monitoring systems employing a send-on-delta method', *Automatica*, 2007, **43**, (2), pp. 332–338

[15] You, K., Xie, L.: 'Kalman filtering with scheduled measurements', *IEEE Trans. Signal Process.*, 2013, **61**, (6), pp. 1520–1530

[16] Wu, J., Jia, Q., Johansson, K., *et al.*: 'Event-based sensor data scheduling: trade-off between communication rate and estimation quality', *IEEE Trans. Autom. Control*, 2013, **58**, (4), pp. 1041–1046

[17] Yang, C., Shi, L.: 'Deterministic sensor data scheduling under limited communication resource', *IEEE Trans. Signal Process.*, 2011, **59**, (10), pp. 5050–5056

[18] Wang, G., Chen, J., Sun, J.: 'Stochastic stability of extended filtering for nonlinear systems with measurement packet losses', *IET Control Theory Appl.*, 2013, **7**, (17), pp. 2048–2055

[19] Chen, B., Zhang, W.A., Yu, L.: 'Distributed finite-horizon fusion Kalman filtering for bandwidth and energy constrained wireless sensor networks', *IEEE Trans. Signal Process.*, 2014, **62**, (4), pp. 797–812

[20] Chen, B., Zhang, W.A., Yu, L., *et al.*: 'Distributed fusion estimation with communication bandwidth constraints', *IEEE Trans. Autom. Control*, 2015, **60**, (5), pp. 1398–1403

[21] Song, H., Zhang, W.A., Yu, L.: 'Hierarchical fusion in clustered sensor networks with asynchronous local estimates', *IEEE Signal Process. Let.*, 2014, **21**, (12), pp. 1506–1510

[22] Wang, G., Chen, J., Sun, J.: 'On sequential Kalman filtering with scheduled measurements'. Proc. 3rd IEEE Intl. Conf. Cyber Technology Automation Control and Intelligent Systems, Nanjing, China, May 2013, pp. 450–455

[23] You, K., Fu, M., Xie, L.: 'Mean square stability for Kalman filtering with Markovian packet losses', *Automatica*, 2011, **47**, (12), pp. 2647–2657

[24] Sinopoli, B., Schenato, L., Franceschetti, M., *et al.*: 'Kalman filtering with intermittent observations'. *IEEE Trans. Autom. Control*, 2004, **49**, (9), pp. 1453–1464

[25] Garone, E., Sinopoli, B., Casavola, A.: 'LQG control for distributed systems over TCP-like erasure channels'. Proc. 48th IEEE Conf. Decision and Control, New Orleans, LA, USA, December 2007, pp. 44–49

[26] Mahalik, N.: '*Sensor networks and configuration: fundamentals, standards, platforms, and applications*' (Springer–Verlag, Berlin Heidelberg, 2007)

[27] Garone, E., Sinopoli, B., Goldsmith, A.J., *et al.*: 'LQG control for MIMO systems over multiple erasure channels with perfect acknowledgment', *IEEE Trans. Autom. Control*, 2012, **57**, (2), pp. 450–456

[28] Mainwaring, A., Culler, D., Polastre, J., *et al.*: 'Wireless sensor networks for habitat monitoring'. Intl. Workshop WSN Applications, Altanlta, GA, USA, September 2002, pp. 88–98

[29] Dey, S., Leong, A., Evans, J.: 'Kalman filtering with faded measurements', *Automatica*, 2008, **45**, (10), pp. 2223–2233

[30] Kotecha, J., Djuric, P.: 'Gaussian particle filtering', *IEEE Trans. Signal Process.*, 2003, **51**, (10), pp. 2592–2601

[31] Leong, A., Dey, S., Nair, G.: 'Quantized filtering schemes for multi-sensor linear state estimation: Stability and performance under high rate quantization', *IEEE Trans. Signal Process.*, 2013, **61**, (15), pp. 3852–3865

[32] Payaró, M., Palomar, D.: 'Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels', *IEEE Trans. Inf. Theory*, 2009, **55**, (8), pp. 3613–3628

9 Appendix

9.1 Proof of Lemma 2

(1) Fact (1) together with Fact (2) is equivalent to showing the minimiser and the minimum value of matrix-valued function $\mathbf{T}_s, \forall s = 1, 2, \ldots, m$, with respect to multiple vector-valued variables $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_s \in \mathbb{R}^{n\times 1}$. For convenience of notation, denote $\mathscr{l}_s = (\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_s)$. We first make extensive use of differential of general matrix-valued function $\mathbf{F}$ with respect to a matrix (vector) argument $\mathbf{X}$ [32].

Definition 1: Let $\mathbf{F}$ be a differentiable $m \times n$ real matrix function of a $p \times q$ matrix of real variables $\mathbf{X}$. The *Jacobian matrix* of $\mathbf{F}$ at $\mathbf{X}$ is given by the $mn \times pq$ matrix

$$\mathrm{D}_{\mathbf{X}}\mathbf{F}(\mathbf{X}) = \frac{\partial\, \mathrm{vec}\mathbf{F}(\mathbf{X})}{\partial(\mathrm{vec}\mathbf{X})^{\mathrm{T}}}.$$

Then vectorising the differential $\mathrm{d}\mathbf{T}_s$ reads that

$$\mathrm{dvec}\mathbf{T}_s = \mathbf{J}_{1,s}\,\mathrm{dvec}\mathbf{l}_1 + \mathbf{J}_{2,s}\,\mathrm{dvec}\mathbf{l}_2 + \ldots + \mathbf{J}_{s,s}\,\mathrm{dvec}\mathbf{l}_s$$

where the Jacobian matrix of $\mathbf{T}_s$ with respect to $\mathbf{l}_i$ is defined as $\mathbf{J}_{i,s} = \mathbf{J}_{i,s}, (\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_s) = \mathbf{D}_{\mathbf{l}_i}\mathbf{T}_s$, $1 \le j \le s$. To make the results more concrete, let us define:

$$\mathbf{G}_{j,j} \triangleq \mathbf{I}_n \otimes \mathbf{I}_n = \mathbf{I}_{n^2}$$

$$\mathbf{G}_{j,t} \triangleq \eta_{j,t}^2 \mathbf{I}_{n^2} + \sum_{i=j+1}^{t} \eta_{i,t}^2 (\mathbf{E}_i \otimes \mathbf{E}_i)\mathbf{G}_{j,i-1}, \quad j+1 \le t \le s.$$

Therefore, after complicated and tedious matrix computations, the Jacobian matrices become (see equation below) where intentionally, $\eta_{s,s}^2$ was not replaced by 1 for the compactness of the structure of $\mathbf{J}_{j,s}$.

Solving $\mathbf{J}_{j,s} = \mathbf{0}$ yields straightforwardly that

$$\mathbf{E}_j\mathbf{T}_{j-1}\mathbf{c}_j + \mathbf{l}_j r_j = \mathbf{0} \Longrightarrow \mathbf{l}^*_{j,s} = -\mathbf{T}_{j-1}\mathbf{c}_j\left(\mathbf{c}_j^{\mathrm{T}}\mathbf{T}_{j-1}\mathbf{c}_j + r_j\right)^{-1} \triangleq \mathbf{l}_j^{\mathbf{X}}$$

where $\mathbf{T}_{j-1} = \mathbf{T}_{j-1}(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_{j-1}^{\mathbf{X}}) = \mathbf{T}_{j-1}^{\mathbf{X}}$.

Then similarly, by solving $\mathbf{J}_{s,s} = \mathbf{0}$, it gives that

$$\mathbf{E}_s\mathbf{T}_{s-1}\mathbf{c}_i + \mathbf{l}_s\mathbf{r}_s = \mathbf{0} \Longrightarrow \mathbf{l}^*_{s,s} = -\mathbf{T}_{s-1}\mathbf{c}_s\left(\mathbf{c}_s^{\mathrm{T}}\mathbf{T}_{s-1}\mathbf{c}_s + r_s\right)^{-1} \triangleq \mathbf{l}_s^{\mathbf{X}}$$

where $\mathbf{T}_{s-1} = \mathbf{T}_{s-1}(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_{s-1}^{\mathbf{X}}) = \mathbf{T}_{s-1}^{\mathbf{X}}$. It should be clearly noted that $\mathbf{l}^*_{j,s} = \mathbf{l}^*_{j,t}$, $\forall t \geq s$, and then plugging $\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_m^{\mathbf{X}}$ into (30) verifies that $\mathbf{M}_m(\mathbf{X}) = \mathbf{T}_m(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_m^{\mathbf{X}})$.

(2) We show this fact by mathematical induction. When $m = 1$, one can easily verify that $\mathbf{l}_1^{\mathbf{X}}$ minimises $\mathbf{T}_1(\mathbf{l}_1, \mathbf{X})$. Suppose now that it holds for $m = k$; that is, the point $(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_k^{\mathbf{X}})$ minimises $\mathbf{T}_k(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_k, \mathbf{X})$. Then for $m = k + 1$,

$$\mathbf{T}_{k+1} = (1-\lambda_{k+1})\mathbf{T}_k + \lambda_{k+1}(\mathbf{E}_{k+1}\mathbf{T}_k\mathbf{E}_{k+1}^{\mathrm{T}} + \mathbf{l}_{k+1}r_{k+1}\mathbf{l}_{k+1}^{\mathrm{T}})$$

and

$$\mathrm{D}_{\mathbf{T}_k}\mathbf{T}_{k+1} = (1-\lambda_{k+1})(\mathbf{I}_n \otimes \mathbf{I}_n) + \lambda_{k+1}(\mathbf{E}_{k+1} \otimes \mathbf{E}_{k+1}) > \mathbf{0}$$

so one necessary condition for point $(\mathbf{l}_1^*, \mathbf{l}_2^*, \ldots, \mathbf{l}_k^*, \mathbf{l}_{k+1}^*)$ minimising $\mathbf{T}_{k+1}$ is that the point should also minimise $\mathbf{T}_k$, or, $(\mathbf{l}_1^*, \mathbf{l}_2^*, \ldots, \mathbf{l}_k^*)$ minimises $\mathscr{T}_k$. Therefore, $(\mathbf{l}_1^*, \mathbf{l}_2^*, \ldots, \mathbf{l}_k^*) = (\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_k^{\mathbf{X}})$, or, $\mathbf{T}_k = \mathbf{T}_k^{\mathbf{X}}$ when minimising $\mathbf{T}_{k+1}$. Given that $\mathbf{l}_{k+1}$ is independent of $\mathbf{T}_k$ and, $\mathbf{T}_k^{\mathbf{X}} > \mathbf{0}, r_{k+1} > 0$, $\mathbf{T}_{k+1}$ is quadratic and convex in $\mathbf{l}_{k+1}$, and therefore, the minimiser for $\mathbf{T}_{k+1}$ can be found by letting

$$\mathrm{D}_{\mathbf{l}_{k+1}}\mathbf{T}_{k+1} = \lambda_{k+1}\left[\left(\mathbf{E}_{k+1}\mathbf{T}_k^X\mathbf{c}_{k+1} + \mathbf{l}_{k+1}r_{k+1}\right) \otimes \mathbf{I}_n + \mathbf{I}_n \otimes \left(\mathbf{E}_{k+1}\mathbf{T}_k^X\mathbf{c}_{k+1} + \mathbf{l}_{k+1}r_{k+1}\right)\right] = \mathbf{0}$$

which leads to the unique solution $\mathbf{l}_{k+1}^{\mathbf{X}} = -\mathbf{T}_k^{\mathbf{X}}\mathbf{c}_{k+1}\left(\mathbf{c}_{k+1}^{\mathrm{T}}\mathbf{T}_k^{\mathbf{X}}\mathbf{c}_{k+1} + r_{k+1}\right)^{-1}$. So $\left(\mathbf{l}_1^{\mathbf{X}}, \ldots, \mathbf{l}_k^{\mathbf{X}}, \mathbf{l}_{k+1}^{\mathbf{X}}\right)$ minimises $\mathbf{T}_{k+1}$. This completes the proof.

(3) Observe that the function $\mathbf{T}_m$ is affine in the variable $\mathbf{X}$. Let $\mathbf{X} \leq \mathbf{Y}$, and it yields that

$$\begin{aligned}\mathbf{M}_m(\mathbf{X}) &= \mathbf{T}_m\left(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_m^{\mathbf{X}}, \mathbf{X}\right)\\ &\overset{(a)}{\leq} \mathbf{T}_m\left(\mathbf{l}_1^{\mathbf{Y}}, \mathbf{l}_2^{\mathbf{Y}}, \ldots, \mathbf{l}_m^{\mathbf{Y}}, \mathbf{X}\right)\\ &\overset{(b)}{\leq} \mathbf{T}_m\left(\mathbf{l}_1^{\mathbf{Y}}, \mathbf{l}_2^{\mathbf{Y}}, \ldots, \mathbf{l}_m^{\mathbf{Y}}, \mathbf{Y}\right)\\ &\overset{(c)}{=} \mathbf{M}_m(\mathbf{Y})\end{aligned}$$

where (a) is because $\mathbf{L}_m^{\mathbf{X}}$ minimises the function $\mathbf{T}_m$ with respect to variables $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m$, then for any $\boldsymbol{\ell}_m \neq \boldsymbol{\ell}_m^{\mathbf{X}}$, say, $\boldsymbol{\ell}_m = \boldsymbol{\ell}_m^{\mathbf{Y}}$, that is, (a) holds true. (b) is due to $\mathbf{T}_m$ is affine in the variable $\mathbf{X}$ and (c) follows from Fact (2).

(4) Let $\mathbf{Z} = \tau\mathbf{X} + (1-\tau)\mathbf{Y}$, where $\tau \in [0, 1]$. Notice that (see equation below) Assume that $\mathbf{M}_s(\mathbf{Z}) \geq \tau\mathbf{M}_s(\mathbf{X}) + (1-\tau)\mathbf{M}_s(\mathbf{Y})$. Then (see equation below) Therefore, the fact holds true.

(5) Note that (see equation below) where $\eta_{0,m}^2 = \prod_{j=1}^m (1-\lambda_j), \mathbf{E}_0 = \mathbf{I}_n, r_0 = 0,$ and $\mathbf{E}_j^{\mathbf{X}}\mathbf{T}_{j-1}(\mathbf{E}_j^{\mathbf{X}})^{\mathrm{T}} + \mathbf{l}_j^{\mathbf{X}}r_j(\mathbf{l}_j^{\mathbf{X}})^{\mathrm{T}} \geq \mathbf{0}$, $1 \leq j \leq m$.

(6) The first inequality follows directly from Fact (5) and linearity of expectation, that is,

$$\mathbb{E}\left[\mathbf{M}_m(\mathbf{X})\right] \geq \prod_{j=1}^m (1-\lambda_j)\mathbb{E}[\mathbf{X}].$$

The second inequality is due to Fact (4) which implies the concavity of the function $\mathbf{M}_m(\mathbf{X})$, and therefore in the light of Jensen's inequality, it readily gives that

$$\mathbf{M}_m(\mathbb{E}[\mathbf{X}]) \geq \mathbb{E}\left[\mathbf{M}_m(\mathbf{X})\right].$$

9.2 Proof of Lemma 3

We only prove Fact (6) because the others can be derived directly from Lemma 2.

(6) According to Fact (7) above, it gives that $\bar{\mathbf{X}} \geq \boldsymbol{\varphi}(\bar{\mathbf{X}}) \geq \prod_{j=1}^m (1-\lambda_j)\mathbf{A}\bar{\mathbf{X}}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$. Since $\left(\mathbf{A}, \mathbf{Q}^{(1/2)}\right)$ is controllable, then there must exist an $\hat{\mathbf{X}} > \mathbf{0}$ subject to the Lyapunov equation $\hat{\mathbf{X}} = \prod_{j=1}^m (1-\lambda_j)\mathbf{A}\hat{\mathbf{X}}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$ if $\sqrt{\prod_{j=1}^m (1-\lambda_j)}\mathbf{A}$ is asymptotically stable. Accordingly, it follows that

$$\bar{\mathbf{X}} - \hat{\mathbf{X}} > \prod_{j=1}^m (1-\lambda_j)\mathbf{A}(\bar{\mathbf{X}} - \hat{\mathbf{X}})\mathbf{A}^{\mathrm{T}}$$

$$\begin{aligned}\mathbf{J}_{j,s} &= \Big[\eta_{j,s}^2\mathbf{I}_{n^2} + \sum_{k=j+1}^{s} \eta_{k,s}^2(\mathbf{E}_k \otimes \mathbf{E}_k)\mathbf{G}_{j,k}\Big]\\ &\quad \times \left[\left(\mathbf{E}_j\mathbf{T}_{j-1}\mathbf{C}_j' + \mathbf{l}_j\mathbf{r}_j\right) \otimes \mathbf{I}_n + \mathbf{I}_n \otimes \left(\mathbf{E}_j\mathbf{T}_{j-1}\mathbf{C}_j' + \mathbf{l}_j\mathbf{r}_j\right)\right],\\ &\quad 1 \leq j \leq s-1\\ \mathbf{J}_{s,s} &= \eta_{s,s}^2\Big[\left(\mathbf{E}_s\mathbf{T}_{s-1}\mathbf{C}_s^{\mathrm{T}} + \mathbf{l}_s\mathbf{r}_s\right) \otimes \mathbf{I}_n\\ &\quad + \mathbf{I}_n \otimes \left(\mathbf{E}_s\mathbf{T}_{s-1}\mathbf{C}_s^{\mathrm{T}} + \mathbf{l}_s\mathbf{r}_s\right)\Big]\end{aligned}$$

$$\begin{aligned}\mathbf{M}_1(\mathbf{Z}) &= \mathbf{T}_1(\mathbf{l}_1^{\mathbf{Z}}, \mathbf{Z})\\ &= \eta_{0,1}^2\mathbf{Z} + \eta_{1,1}^2\Big[(\mathbf{I}_n + \mathbf{l}_1^{\mathbf{Z}}\mathbf{c}_1^{\mathrm{T}})\mathbf{Z}(\mathbf{I}_n + \mathbf{l}_1^{\mathbf{Z}}\mathbf{c}_1^{\mathrm{T}})^{\mathrm{T}}\\ &\quad + \tau\mathbf{l}_1^{\mathbf{Z}}r_1(\mathbf{l}_1^{\mathbf{Z}})^{\mathrm{T}} + (1-\tau)\mathbf{l}_1^{\mathbf{Z}}r_1(\mathbf{l}_1^{\mathbf{Z}})^{\mathrm{T}}\Big]\\ &= \tau\Big[\eta_{0,1}^2\mathbf{X} + \eta_{1,1}^2((\mathbf{I}_n + \mathbf{l}_1^{\mathbf{Z}}\mathbf{c}_1^{\mathrm{T}})\mathbf{X}(\mathbf{I}_n + \mathbf{l}_1^{\mathbf{Z}}\mathbf{c}_1^{\mathrm{T}})^{\mathrm{T}} + \mathbf{l}_1^{\mathbf{Z}}r_1(\mathbf{l}_1^{\mathbf{Z}})^{\mathrm{T}})\Big]\\ &\quad + (1-\tau)\Big[\eta_{0,1}^2\mathbf{Y} + \eta_{1,1}^2\left((\mathbf{I}_n + \mathbf{l}_1^{\mathbf{Z}}\mathbf{c}_1^{\mathrm{T}})\mathbf{Y}(\mathbf{I}_n + \mathbf{l}_1^{\mathbf{Z}}\mathbf{c}_1^{\mathrm{T}})^{\mathrm{T}} + \mathbf{l}_1^{\mathbf{Z}}r_1(\mathbf{l}_1^{\mathbf{Z}})^{\mathrm{T}}\right)\Big]\\ &= \tau\mathbf{T}_1(\mathbf{l}_1^{\mathbf{Z}}, \mathbf{X}) + (1-\tau)\mathbf{T}_1(\mathbf{l}_1^{\mathbf{Z}}, \mathbf{Y})\\ &\geq \tau\mathbf{T}_1(\mathbf{l}_1^{\mathbf{X}}, \mathbf{X}) + (1-\tau)\mathbf{T}_1(\mathbf{l}_1^{\mathbf{Y}}, \mathbf{Y})\\ &= \tau\mathbf{M}_1(\mathbf{X}) + (1-\tau)\mathbf{M}_1(\mathbf{Y}).\end{aligned}$$

implying there exists a $\hat{\mathbf{Q}} > \mathbf{0}$ such that

$$\bar{\mathbf{X}} - \hat{\mathbf{X}} = \prod_{j=1}^{m} (1-\lambda_j)\mathbf{A}(\bar{\mathbf{X}} - \hat{\mathbf{X}})\mathbf{A}^{\mathrm{T}} + \hat{\mathbf{Q}}.$$

Thus, $\bar{\mathbf{X}} - \hat{\mathbf{X}} > \mathbf{0}$, or $\bar{\mathbf{X}} > \hat{\mathbf{X}} > \mathbf{0}$. This completes the proof.

9.3 Proof of Lemma 4

(1) Note that $\mathbf{L}_m(\mathbf{Y})$ is affine in $\mathbf{Y}$ and $\mathbf{L}_m(\mathbf{Y}) \geq \mathbf{0}, \forall \mathbf{Y} \geq \mathbf{0}$, and $\mathbf{L}_m(\mathbf{Y}) \geq \mathbf{L}_m(\mathbf{Z})$, for $\mathbf{Y} \geq \mathbf{Z}$. There exist constants $0 \leq r < 1$ and $t \geq 0$ such that $\mathbf{L}_m(\bar{\mathbf{Y}}) \leq r\bar{\mathbf{Y}} < \bar{\mathbf{Y}}$ and $\mathbf{W} \leq t\bar{\mathbf{Y}}$, respectively. Then

$$0 \leq \mathbf{L}_m^k(\mathbf{W}) \leq t\mathbf{L}_m^k(\bar{\mathbf{Y}}) \leq \mathrm{tr}^k\bar{\mathbf{Y}}. \tag{36}$$

Therefore, we have $\mathbf{0} \leq \lim_{k\to\infty} \mathscr{L}_m^k(\mathbf{W}) \leq \lim_{k\to\infty} \mathrm{tr}^k\bar{\mathbf{Y}} \to \mathbf{0}$ given that $0 \leq r < 1$.

(2) Based on (36) above, for any initialisation $\mathbf{Y}_0 \geq \mathbf{0}$ and any $\mathbf{U} > \mathbf{0}$, there always exist two constants $t_{\mathbf{Y}_0} \geq 0$ and $t_{\mathbf{U}} \geq 0$ such that $\mathbf{Y}_0 \leq t_{\mathbf{Y}_0}\bar{\mathbf{Y}}$ and $\mathbf{U} \leq t_{\mathbf{U}}\bar{\mathbf{Y}}$, which are independent of k. Therefore, similar arguments in (36) lead to

$$\begin{aligned}\mathbf{Y}_k = \mathbf{L}_m^k(\mathbf{Y}_0) + \sum_{s=0}^{k-1}\mathbf{L}_m^s(\mathbf{U}) &\leq t_{\mathbf{Y}_0}r^k\bar{\mathbf{Y}} + \sum_{s=0}^{k-1} t_{\mathbf{U}}r^s\bar{\mathbf{Y}} \\ &= \left(t_{\mathbf{Y}_0}r^k + t_{\mathbf{U}}\frac{1-r^k}{1-r}\right)\bar{\mathbf{Y}}.\end{aligned}$$

Thus, the result on boundedness of the sequence Y_k holds true.

9.4 Proof of Lemma 5

Observe that $\boldsymbol{\phi}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{Y}) = \mathbf{L}_m(\mathbf{Y}) + \mathbf{Q} + \mathbf{N}_m$, where $\mathbf{N}_m := \sum_{j=0}^{m}\eta_{j,m}^2\left(\mathbf{E}_j\mathbf{N}_{j-1}\mathbf{E}_j' + \mathbf{l}_j r_j \mathbf{l}_j^{\mathrm{T}}\right) \geq \mathbf{0}$ with $\mathbf{N}_0 = \mathbf{0}$, $\mathbf{Q} \geq \mathbf{0}$, and $r_j \geq 0, j = 0, 1, \ldots, m$. Therefore,

$$\bar{\mathbf{P}} > \boldsymbol{\phi}_m(\bar{\mathbf{l}}_1, \bar{\mathbf{l}}_2, \ldots, \bar{\mathbf{l}}_m, \bar{\mathbf{P}}) = \mathbf{L}_m(\bar{\mathbf{P}}) + \mathbf{Q} + \mathbf{N}_m \geq \mathbf{L}_m(\bar{\mathbf{P}}).$$

That is, $\bar{\mathbf{P}} > \mathbf{L}_m(\bar{\mathbf{P}})$. Thus, $\mathbf{L}_m(\mathbf{Y})$ satisfies the condition of Lemma 4. Considering definition of $\boldsymbol{\varphi}(\mathbf{P}_k)$, it yields that

$$\begin{aligned}\mathbf{P}_{k+1} = \boldsymbol{\varphi}(\mathbf{P}_k) &\leq \boldsymbol{\phi}_m(\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_m, \mathbf{P}_k) \\ &= \mathbf{L}_m(\mathbf{P}_k) + \mathbf{Q} + \mathbf{N}_m \\ &= \mathbf{L}_m(\mathbf{P}_k) + \mathbf{U}\end{aligned}$$

where $\mathbf{U} := \mathbf{Q} + \mathbf{N}_m \geq \mathbf{0}$. Then based on fact 2) in Lemma 4, the sequence $\mathbf{P}_k$ is bounded for any $k \geq 0$.

9.5 Proof of Lemma 6

The three statements can be similarly proved by mathematical induction. Thus, due to page limitation, we here only prove the first one. Since $\mathbf{Y}_1 \geq \mathbf{Y}_0$, then the first statement is true for $k = 0$. Then assume that $\mathbf{Y}_{t+1} \geq \mathbf{Y}_t$ holds, so $\mathbf{Y}_{t+2} = f(\mathbf{Y}_{t+1}) \geq f(\mathbf{Y}_t) = \mathbf{Y}_{t+1}$ holds owing to the monotonicity of function $f(\mathbf{Y})$.

$$\begin{aligned}\mathbf{M}_{s+1}(\mathbf{Z}) &= \mathbf{T}_{s+1}(\mathbf{l}_1^{\mathbf{Z}}, \mathbf{l}_2^{\mathbf{Z}}, \ldots, \mathbf{l}_{s+1}^{\mathbf{Z}}, \mathbf{Z}) \\ &= (1-\lambda_{s+1})\mathbf{M}_s(\mathbf{Z}) + \lambda_{s+1} \\ &\quad \times \left[(\mathbf{I}_n + \mathbf{l}_{s+1}^{\mathbf{Z}}\mathbf{c}_{s+1}^{\mathrm{T}})\mathbf{M}_s(\mathbf{Z})(\mathbf{I}_n + \mathbf{l}_{s+1}^{\mathbf{Z}}\mathbf{c}_{s+1}^{\mathrm{T}})^{\mathrm{T}} + \mathbf{l}_{s+1}^{\mathbf{Z}} r_{s+1}(\mathbf{l}_{s+1}^{\mathbf{Z}})^{\mathrm{T}}\right] \\ &\geq (1-\lambda_{s+1})\left[\tau\mathbf{M}_s(\mathbf{X}) + (1-\tau)\mathbf{M}_s(\mathbf{Y})\right] \\ &\quad + \lambda_{s+1}\left[(\mathbf{I}_n + \mathbf{l}_{s+1}^{\mathbf{Z}}\mathbf{c}_{s+1}^{\mathrm{T}})(\mathbf{M}_s(\mathbf{X})\right. \\ &\quad + (1-\tau)\mathbf{M}_s(\mathbf{Y}))(\mathbf{I}_n + \mathbf{l}_{s+1}^{\mathbf{Z}}\mathbf{c}_{s+1}^{\mathrm{T}})^{\mathrm{T}} \\ &\quad \left. + (\tau + 1 - \tau)\mathbf{l}_{s+1}^{\mathbf{Z}} r_{s+1}(\mathbf{l}_{s+1}^{\mathbf{Z}})^{\mathrm{T}}\right] \\ &= \tau\mathbf{T}_{s+1}(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_s^{\mathbf{X}}, \mathbf{l}_{s+1}^{\mathbf{Z}}, \mathbf{X}) \\ &\quad + (1-\tau)\mathbf{T}_{s+1}(\mathbf{l}_1^{\mathbf{Y}}, L_2^{\mathbf{Y}}, \ldots, \mathbf{l}_s^{\mathbf{Y}}, \mathbf{l}_{s+1}^{\mathbf{Z}}, \mathbf{Y}) \\ &\geq \tau\mathbf{T}_{s+1}(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_s^{\mathbf{X}}, \mathbf{l}_{s+1}^{\mathbf{X}}, \mathbf{X}) \\ &\quad + (1-\tau)\mathbf{T}_{s+1}(\mathbf{l}_1^{\mathbf{Y}}, \mathbf{l}_2^{\mathbf{Y}}, \ldots, \mathbf{l}_s^{\mathbf{Y}}, \mathbf{l}_{s+1}^{\mathbf{Y}}, \mathbf{Y}) \\ &\geq \tau\mathbf{M}_{s+1}(\mathbf{X}) + (1-\tau)\mathbf{M}_{s+1}(\mathbf{Y}).\end{aligned}$$

$$\begin{aligned}\mathbf{M}_m(\mathbf{X}) &= \mathbf{T}_m\left(\mathbf{l}_1^{\mathbf{X}}, \mathbf{l}_2^{\mathbf{X}}, \ldots, \mathbf{l}_m^{\mathbf{X}}, \mathbf{X}\right) \\ &= \eta_{0,m}^2\left(\mathbf{E}_0\mathbf{X}\mathbf{E}_0^{\mathrm{T}} + \mathbf{l}_0 r_0 \mathbf{l}_0^{\mathrm{T}}\right) \\ &\quad + \sum_{j=1}^{m}\eta_{j,m}^2\left[\mathbf{E}_j^{\mathbf{X}}\mathbf{T}_{j-1}(\mathbf{E}_j^{\mathbf{X}})^{\mathrm{T}} + \mathbf{l}_j^{\mathbf{X}} r_j (\mathbf{l}_j^{\mathbf{X}})^{\mathrm{T}}\right] \\ &\geq \prod_{j=1}^{m}(1-\lambda_j)\mathbf{X}\end{aligned}$$