State Estimation and Data Fusion With Data-driven Communication

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Engineering and Applied Science
Electrical Engineering

by

Xiaolei Bian

May 2023

Table of Contents

Abstract

This dissertation focuses on the systemic design of proper estimation as well as fusion techniques and data-driven communication schemes to infer the state of a dynamic discrete-time linear system over a wireless network. The goal is to effectively extract and share key information over networks to enhance estimates of system states. Such co-design research is drawn by the need to leverage limited resources (communication bandwidth, energy and computational power) for a multitude of data and data sources in networked systems.

We first build the overarching structure by synthesizing communication and estimator/fuser, and research the collective behaviors of the system.

In the design phase, we jointly construct the data-driven communication scheme and the estimator/fuser based on a single sensor, centralized and distributed estimation fusion architecture, respectively.

Based on this structure, we design data-driven communication schemes both based on measurement innovation and estimation innovation. Moreover, we propose a new metric based on cumulative estimate innovation to smartly pick key information. We derive the corresponding (approximate) minimum mean square error (MMSE) estimators/fusers and the MMSE-optimal weighted least square (WLS) fusers in the closed-form representations. Further, our fusers have guaranteed stability.

These data-driven communication schemes take into account multiple network structures and the interaction between modules of communication and estimation. They utilize the importance of innovation for estimation and can achieve a trade-off between communication costs and estimation performance. The proposed estimators and fusers optimally use information in the triggering decisions to boost estimation performance.

Further, with the aim of optimally calibrating the tradeoff between estimation quality and communication expenses, we treat limited communication resources quantitatively in the joint design and formulate the problem in the framework of optimal estimation by incorporating these limited

resources. Through constructing an auxiliary state vector, the optimization problem on the expected total discounted cost—including estimation error and weighted communication cost—over the infinite horizon is shown to be representable as a Markov Decision Process (MDP) problem. An iterative algorithm is proposed to find the optimal cost and optimal policy. The optimal policy has a peculiar degree-of-freedom reduction property.

Besides all these theoretical contributions, our work carries weight of practical applications. The proposed data-driven communication schemes under our metric have easy setup and can seamlessly integrate different dynamic systems, because the optimal decision policy is algorithmically attainable. The stability study of estimators and fusers ensures their applicability in practical control systems. The designed data-driven communication schemes and estimators/fusers can provide optimal solutions for state estimation over sensor networks; also they are reliable and easy to implement across multiple sensor network architectures.

Keywords: Estimate fusion, data-driven communication, event-trigger, innovation-based communication, MMSE estimator, LS fuser, MDP, dynamic programming, stability

List of Tables

List of Figures

List of Acronyms

BLUE	Best Linear Unbiased Estimator
DDC	Data-Driven Communication
DP	Dynamic Programming
GCSN	Generalized Closed Skew Normal
HMM	Hidden Markov Model
JDE	Joint Decision and Estimation
KL	Kullback-Leibler
LQG	Linear-Quadratic-Gaussian
LS	Least Square
MAP	Maximum A Posteriori
MDP	Markov Decision Process
MLE	Maximum Likelihood Estimator
MMSE	Minimum Mean Square Error
SOA	Send-On-Area
SOD	Send-On-Delta
WLS	Weighted Linear Square

List of Notations

$\mathbb{R}$	the set of real numbers
$\mathbb{Z}$	the set of non-negative integers
$\mathbb{N}$	the set of natural numbers
$\{A\}_0$	the matrix obtained by deleting all 0 rows from the matrix A
$B \succ 0$ $(\succeq 0)$	a positive definitive (positive semi-definite) matrix $B \in \mathbb{R}^{n \times n}$
$\mathcal{N}(\mu, \Gamma)$	the Gaussian distribution with mean μ and covariance matrix $\Gamma \succ 0$
χ_n^2	the chi-square distribution with n degrees of freedom
$\{\zeta\}_0^k$	the set $\{\zeta_0, \cdots, \zeta_k\}$
$f_{\mathrm{x}}(x)$	the probability density function (pdf) of the random vector x
$f_{\mathrm{x}\vert\mathrm{y}}(x\vert y)$	the pdf of a random vector x conditioned on the random vector y
$f_{\mathrm{x},\mathrm{y}}(x, y)$	the joint pdf of random vectors x and y
$\mathbb{E}[\cdot]$	the expectation
$\mathbb{E}^f[\cdot]$	the expectation, where the underlying probability measure is parameterized by the policy f
$\Pr(\cdot)$	the probability
$\Pr^f(\cdot)$	the probability, where the underlying probability measure is parameterized by the policy f
$\vert\vert \cdot \vert\vert$	the Euclidean norm of a vector or the induced 2-norm of a matrix
$\vert\vert \cdot \vert\vert_\infty$	the infinity-norm of a vector
$\vert\vert \cdot \vert\vert_{\sigma_1}$	the Schatten 1-norm
$\vert\vert \cdot \vert\vert_{\sigma_\infty}$	the Schatten ∞-norm
$\lambda_{\max}(\cdot)$	the maximum eigenvalue of a matrix
$\sigma_{\max}(\cdot)$	the largest singular value of a matrix
$\mathrm{tr}(\cdot)$	the trace of a matrix
δ_{ij}	the Kronecker delta
$\mathbb{1}(A)$	the indicator function of a subset A

CHAPTER 1
INTRODUCTION

1.1 Background and motivations

The unbridled aspiration to master facts and laws of nature, and the insatiable appetite to create and rebuild the physical and virtual world through engineering means are among the primal impetuses for human progression. These desires and ambitions have always been compromised throughout the endeavors, prompting humans to surrender to the doctrine that no object can be known with absolute certainty, by any means. Exact reasoning behind this mentality remains an elusive topic of debate: whether, psychologically, people are humbled by deficiencies in resources and tools at their disposal; or, philosophically, absolute truths are prohibited from being understood, or unattainable, or nonexistent by some basic law of nature. While never shrewd enough to resolve such subject matter, humans have often been viable enough to find wriggle room. Learning to cohabit with uncertainties, after all, still leads to ample knowledge that meets both scientific and applied ends.

Scientists and engineers spare no effort in bringing about an orderly and intelligent world interwoven both physically and virtually. Networked systems are manifesting examples of the progress being made. Such state-of-the-art architectures include Industrial Cyber-Physical Systems, Industrial Internet of Things, and Networked Control Systems [1–3]. A networked system usually features a high degree of decentralized design, in which great autonomy is delegated to its individual nodes. In such a system, the dichotomy between the vassal units at the low level and the managerial coordinations and controls at the high level ceases to be valid. Due to its decentralized and fluid nature, the interplay between various functional modules reverberates through the entire system simultaneously and ubiquitously, necessitating a cohesive treatment of the collective behaviors of the system.

It is important to note that networked systems are strained by limited resources, the most press-ing ones being energy, bandwidth and computational power. With resources-awareness in mind, engineers are developing systems that meet taxing demands through intelligent allocation of resources.

Knowledge on some object of interest in scientific and engineering terms are typically numerical values. Inferring numerical value of something by uncertain and indirect observations is the process of estimation [4]. Such object is often the state of a stochastic dynamical system. Not only is state estimation important in its own right, e.g., positions and velocities of an object in navigation, but also it forms an inalienable part in control theory.

The aforementioned numerical values—collectively known as data—are generated by individual sensors. In networked systems, multi-sensor configurations are often applied in lieu of single-sensor ones (i.e. sensors at one location), since they reveal more information. However, the presence of multiple data sources also adds to the complexity of the systems. Data fusion is called up to integrate data from multiple sources to produce more consistent, accurate, and/or useful information than what would otherwise be provided by data from any individual source [5]. In particular, the data fusion for estimation, commonly referred to as estimation fusion, focuses on estimating the parameters or states using data from multiple sensors. Any fusion process requires certain amount of computational power, which in turn costs energy as well.

No development of a networked system could have come to fruition without extensive research into befitting communication schemes. Communication is the transmission of information from one point to another through a succession of processes [6]. Communication in a networked system may be bottlenecked by bandwidth or energy; also, it may waste or overdrive the computational power in a fusion center by overloading the data.

Treading a fine line between reducing uncertainties and conserving resources, the research on data-driven communication (DDC) schemes prospers. A DDC scheme allocates energy and bandwidth efficiently and decentralizes the computational overload to local sensors.

In short, the joint study of state estimation, estimation fusion and DDC are ratcheting up momentum in that it fosters holistic approaches to deliver more precise estimates at lower cost of resources. This dissertation focuses on systemically researching estimation as well as fusion techniques and DDC schemes to infer states of a discrete-time dynamical system over networks.

1.2 Preliminaries of optimal estimation and estimation fusion

In this section, we provide preliminaries of optimal estimation, in particular minimum mean square error (MMSE) estimation and Least square (LS) estimation, needed for our research. Then we delve into estimation fusion. The estimators and fusers in this dissertation are all constructed in the spirit of MMSE and LS.

Optimal estimation is an algorithmic process that treats observations (measurements) to yield an estimate of a variable or parameter of interest, which optimizes a certain criterion [7]. Under different criteria of optimality, various estimator have been developed, for example, the LS estimator, the MMSE estimator, the maximum likelihood (MLE) estimator and the maximum a posteriori (MAP) estimator [4]. Among them, the LS estimator, the MMSE estimator and the MLE estimator are three prominent ones [8, 9].

1.2.1 Bayesian approach, MMSE estimation and Kalman filter

Bayesian approach

The object of interest is a random variable x with known prior distribution $f_{\mathbf{x}}(x)$. Conceptually, one wishes to improve the understanding of the random variable from its prior knowledge with an inaccurate measurement y of x, using Bayes' formula [4]:

$$f_{\mathbf{x}|\mathbf{y}}(x|y) = \frac{f_{\mathbf{y}|\mathbf{x}}(y|x) f_{\mathbf{x}}(x)}{f_{\mathbf{y}}(y)} \tag{1.1}$$

where $f_{\mathbf{x}|\mathbf{y}}(x|y)$ is the posterior pdf.

The Hidden Markov Model (HMM) in Fig. 1.1 (adapted from [10]) is the simplest framework to show how recursive Bayesian estimation, known as Bayes filter, pins down our best possible knowledge of the random variable in successive steps. The dynamical system evolves in discrete

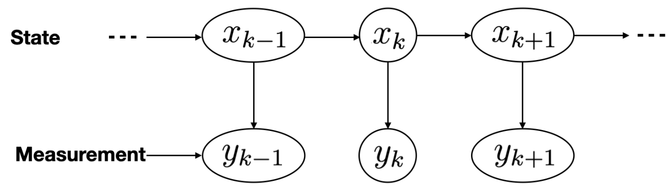


Figure 1.1: Hidden Markov Model

time and explicit state information is hidden, but the system communicates a measurement of the state at each timestamp. We model the state x_k and measurement y_k ($k \in \mathbb{Z}$) as random variables. This model satisfies the assumption of completeness: the system being Markovian, which enables transition model, and the current measurement depending only on the current state, known as measurement model [10]. The paradigm of the Bayes filter is shown in Fig. 1.2.

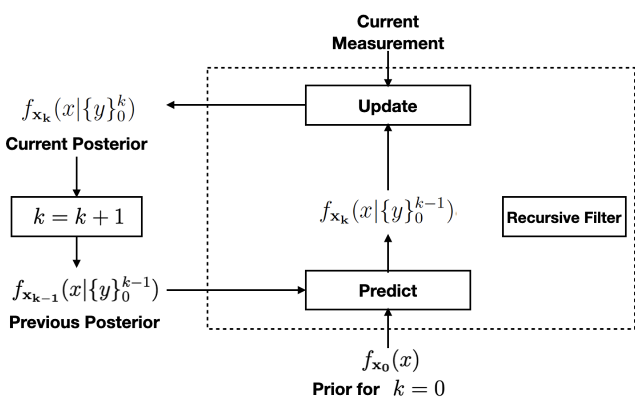


Figure 1.2: Paradigm of the Bayes filter

Bayes filter recursively estimates the system state that was inaccurately measured with noise. Each iteration refines its knowledge of the system state in two consecutive steps.

1) Prediction step

Based on a known state posterior of system state at time $k-1$, $f_{\mathbf{x_{k-1}}}(x|\{y\}_0^k)$, along with the transition model $f_{\mathbf{x_k}}(x|x_{k-1})$, the best guess is made to infer the state posterior $f_{\mathbf{x_k}}(x|\{y\}_0^k)$ at time k:

$$f_{\mathbf{x_{k-1}}}(x|\{y\}_0^k) = \int_{\mathbb{R}^n} f_{\mathbf{x_k}}(x|x_{k-1}) f_{\mathbf{x_{k-1}}}(x|\{y\}_0^{k-1}) dx_{k-1} \tag{1.2}$$

2) Update step

The measurement at time k emerges as an additional piece of information available that improves the knowledge and confidence on top of the predicted belief. $f_{\mathbf{y_k}}(y|x_k)$ is the measurement likelihood and provides information on the probability density on resolving $\mathbf{y_k}$ at y.

$$f_{\mathbf{x_k}}(x|\{y\}_0^k) = \frac{f_{\mathbf{y_k}}(y|x_k) f_{\mathbf{x_k}}(x|\{y\}_0^{k-1})}{f_{\mathbf{y_k}}(y|\{y\}_0^{k-1})} \tag{1.3}$$

Obviously, the updated state posterior leverages how well the current measurement is explained by the current system state, on top of the best guess of state posterior from the prediction step based on all information available barring current measurement.

MMSE estimator

The MMSE estimator is one that minimizes the expected value of the squared difference between what's to be estimated x and its estimate $\hat{x}$, and is given by

$$\hat{x}^{\text{MMSE}} = \underset{\hat{x}(y)}{\operatorname{argmin}} E[(x - \hat{x}(y))'(x - \hat{x}(y))] \tag{1.4}$$

where y is the observation of x. It is an adaptation of Bayesian approach with quadratic cost function.

Consider the follow general discrete-time stochastic system

$$x_{k+1} = f_k(x_k, w_k) \tag{1.5}$$

$$y_k = h_k(x_k, v_k), k \in \mathbb{Z} \tag{1.6}$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ are the system state and measurement, respectively. The process noise w_k and the measurement noise v_k are white, mutually independent and independent of the past and the current state. They are also mutually independent for any time k. The MMSE estimator at time k consists of two steps.

1) Time update

$$\hat{x}_{k|k-1} = E[x_k|\{y\}_0^{k-1}] \tag{1.7}$$

$$= \int_{\mathbb{R}^n} x f_{\mathbf{x_k}}(x|\{y\}_0^{k-1})dx \tag{1.8}$$

$$P_{k|k-1} = E[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})'|\{y\}_0^{k-1}]$$

$$= \int_{\mathbb{R}^n} (x - \hat{x}_{k|k-1})(x - \hat{x}_{k|k-1})' f_{\mathbf{x_k}}(x|\{y\}_0^{k-1})dx. \tag{1.9}$$

2) Measurement update

$$f_{\mathbf{x_k}}(x|\{y\}_0^k) = \frac{f_{\mathbf{x_k}}(x|\{y\}_0^{k-1}) f_{\mathbf{y_k}}(y|\{y\}_0^{k-1}, x_k)}{f_{\mathbf{y_k}}(y|\{y\}_0^{k-1})} \tag{1.10}$$

$$\hat{x}_k = E[x_k|\{y\}_0^k] = \int_{\mathbb{R}^n} x f_{\mathbf{x_k}}(x|\{y\}_0^k)dx \tag{1.11}$$

$$P_k = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)'|\{y\}_0^k]$$

$$= \int_{\mathbb{R}^n} (x - \hat{x}_k)(x - \hat{x}_k)' f_{\mathbf{x_k}}(x|\{y\}_0^k)dx. \tag{1.12}$$

In practice, an adaptation for measurement model (1.6) is applied when measurement y_k belongs to a region R_k^O at time k, where $R_k^O \in \mathbb{R}^m$ is a subset of the measurement space. For example, the measurement may be quantized, voluntarily dropped due to limited information content, or, missing as a result of being outside observation spectrum. In both cases, the fact that data are quantized/dropped/missing at time k carries useful information about the state.

Let $M_k = \{y_k \in R_k^O\}$, we have

$$f_{\mathbf{x_k}}(x|\{y\}_0^k) = \frac{(\int_{M_k} f_{\mathbf{y_k}|\mathbf{x_k}}(y|x)dy)) f_{\mathbf{x_k}}(x|\{y\}_0^{k-1})}{\int (\int_{M_k} f_{\mathbf{y_k}|\mathbf{x_k}}(y'|x')dy')) f_{\mathbf{x_k}}(x'|\{y\}_0^{k-1})dx'}. \tag{1.13}$$

The MMSE estimation can be applied to a variety of estimation problems, including linear and nonlinear systems, discrete- and continuous-time systems, and one- and multi-dimensional systems. Clear update of the condition mean may be impossible or complicated. As a result, various approximation techniques have been developed.

Kalman filter

The Kalman filter is the optimal MMSE state estimator under the linear-Gaussian assumption [4].

Consider the following discrete-time stochastic linear time-invariant system

$$x_{k+1} = Ax_k + w_k \tag{1.14}$$

$$y_k = Cx_k + v_k, k \in \mathbb{Z} \tag{1.15}$$

where $x_k \in \mathbb{R}^n$ is the system state, $A \in \mathbb{R}^{n\times n}$ ($n \in \mathbb{N}$) is the system matrix, $y_k \in \mathbb{R}^m$ ($m \in \mathbb{N}$) is the measurement, $C \in \mathbb{R}^{m\times n}$ is the output matrix, w is zero-mean Gaussian white process noise with covariance Q, and v is zero-mean Gaussian white measurement noise with covariance R. The initial state is $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$. We further assume that x_0, w_k and v_k are mutually uncorrelated for any time k. The Kalman filter is

$$\hat{x}_{k|k-1} = \mathbb{E}[x_k|\{y\}_0^{k-1}] = A\hat{x}_{k-1} \tag{1.16}$$

$$P_{k|k-1} = \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})'] = AP_{k-1}A' + Q \tag{1.17}$$

$$\hat{x}_k = \mathbb{E}[x_k|\{y\}_0^k] = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}) \tag{1.18}$$

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)'] = P_{k|k-1} - K_kCP_{k|k-1} \tag{1.19}$$

$$K_k = P_{k|k-1}C'(CP_{k|k-1}C' + R)^{-1}. \tag{1.20}$$

Consequently, based on the previous estimate $(\hat{x}_{k-1}, P_{k-1})$, the estimation can be written compactly as

$$(\hat{x}_{k|k-1}, P_{k|k-1}, \hat{x}_k, P_k, K_k) = \mathbf{KF}(A, C, Q, R, y_k). \tag{1.21}$$

1.2.2 LS estimator

The LS estimator, dating back to the early 19th century with the work of Legendre and Gauss, is the oldest and one of the simplest methods for estimation [11]. The basic idea of the LS method is to choose a parameter value of a model that fits best the available data such that the sum of squares of the fitting errors is least. Consider a linear observation model for x with an additive error v_i [4]

$$z_i = G_i x + v_i, \quad i = 1, 2, \cdots, n. \tag{1.22}$$

The **linear weighted LS (WLS)** estimator of x approximates x with the smallest fitting error:

$$\hat{x}^{\text{LS}} = \text{arg}\min_{\hat{x}}[z^n - H^n \hat{x})]' W^n [z^n - H^n \hat{x}] \tag{1.23}$$

where z^n is the stack of data z_i, v^n the stack of v_i, H^n the stack of the linear fitting model to data z_i and W^n a symmetric positive weighting matrix. If v^n has zero mean and covariance R^n, the MMSE-optimal WLS estimator is

$$\hat{x}^{\text{LS}} = P_n^{\text{min}} (H^n)'(R^n)^{-1} z^n \tag{1.24}$$

$$P_n^{\text{min}} = [(H^n)'(R^n)^{-1} H^n]^{-1}. \tag{1.25}$$

This WLS estimator minimizes MSE of all linear unbiased estimators using a linear data model of a nonrandom estimand or a random estimand with unknown prior mean. The best linear unbiased estimator (BLUE) is a linear estimator that is unbiased and best among all linear estimators under a certain optimality criterion. Note that linear LS estimator (W^n being an identity matrix) is BLUE for fitting error.

1.2.3 Estimation fusion

Estimation fusion, also known as data fusion for estimation, involves combining information from multiple data sources—sensors usually—to achieve a more complete and accurate understanding of a physical system or process [12]. Estimation fusion is used in a wide range of applications such as robotics, autonomous vehicles, navigation systems, and tracking systems. In these applications, multiple sensors such as cameras, lidars, radars, and GPS are used to collect data and estimate the state of the system or object being tracked [13]. However, each sensor has its own limitations and uncertainties, which can lead to inaccurate or incomplete estimates.

Three fundamental estimation fusion architectures are in use: centralized fusion, distributed fusion and hybrid fusion, depending on whether raw measurements are processed at the local sensors [14, 15]. The architectures of them are shown in Fig 1.2, 1.3 and 1.4, respectively.

In centralized fusion, raw data from multiple sensors is transmitted to a fusion center and processed. It requires no computational power for individual sensors and provides best estimation performance in theory. But it also poses high demands on the bandwidth of communication channels and processing power of the fusion center.

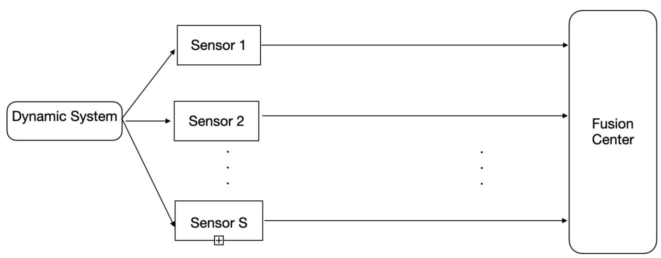


Figure 1.3: Architecture of centralized fusion

In distributed fusion, each sensor processes the local measurements and generates a local estimate of the system state. The local estimates are then communicated to a fusion center, and the fusion algorithms are in place to combine the estimates from different sensors and generate a global estimate of the system state. Distributed fusion has low requirements on channel bandwidth and central processing capacity. In the meantime, it serves as a good candidate for unreliable/limited communication channels since each local estimate sums up all information about previous measurements. In summary, it manifests itself as a robust and scalable solution compared to centralized fusion and is well suited for applications that require high viability and scalability, such as wireless sensor networks and distributed control systems. It is also more challenging since its construct is of more sophisticated nature.

In the absence of a fusion center, each node communicates with its neighbors (peer-to-peer network). To solve the distributed estimation fusion problem over a peer-to-peer network, the community has developed a number of approaches, including gossip, diffusion, and consensus algorithms. Among them, consensus algorithm is widely used.

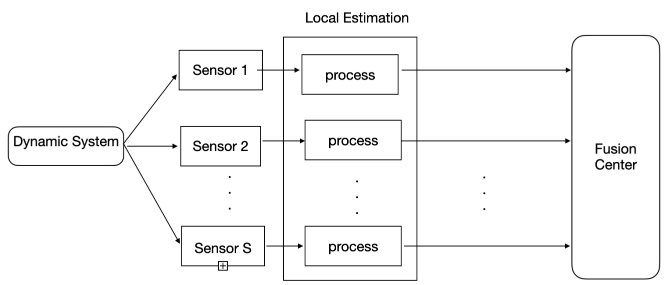


Figure 1.4: Architecture of distributed fusion

Hybrid fusion is a synthesis of the centralized and distributed fusion mechanisms. The fusion center may receive either raw measurements or processed ones—i.e., local estimates.

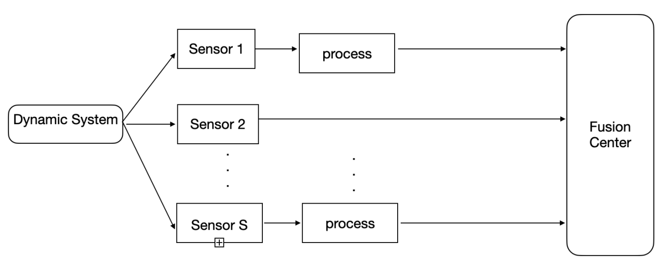


Figure 1.5: Architecture of hybrid fusion

1.3 Survey of data-driven communication and its applications in state estimation and fusion

DDC is closely related to the "event-based/adaptive/Lebesgue sampling" [16–19]. The idea can date back, at the earliest, to 1959, when Ellis noticed that "The most suitable sampling is by transmission of only significant data, as the new value obtained when the data are changed by a given increment" [20]. Consequently, "it is not necessary to sample periodically, but only when quantized data change from one possible value to the next" [20].

Recent years witnesses a revival of these topics motivated by the growing popularity of systems over networks. Such systems are constricted by limited resources. In fact, experimental studies show that communication is a major source of energy consumption in sensors. A naive way to mitigate these issues is to simply reduce the communication rate, but this would inevitably lead to degraded estimation quality. This drawback has invited the development of a broad range of effective DDC strategies [21, 22], where data are transmitted only when their significance in some measure exceeds a specified threshold, as determined by the data quality—not by time. Such strategies provide an opportunity to do good estimation using reduced transmission since they allocate resources to important data. Define γ_k as a binary variable for a sensor, where $\gamma_k = 1$ indicates that sensor transmits data at time k and $\gamma_k = 0$ otherwise. Then, the design of a DDC scheme in a discrete-time system can be modeled as a decision process

$$\gamma_k = \begin{cases} 1 & \text{if } \boldsymbol{\mu}(Y) \leq \delta_k \\ 0 & \text{otherwise,} \end{cases} \tag{1.26}$$

where $\boldsymbol{\mu}(\cdot)$ is a metric of data Y and δ_k a threshold.

Note that we use "event-triggered" and "data-driven" schemes interchangeably throughout this dissertation. They are closely associated yet not identical terms. The event-triggered communication schemes command broader sense in that they are touched off by conditions, which include, but are not limited to, data. Communication might be triggered by a change in a hybrid/switched system or an occurrence of target maneuver/malfunctioning sensor. The cases considered in this dissertation can be subsumed into the data-driven category, and therefore, it largely defers to personal tastes to pick the term for use.

Mainstream literature on data-driven estimators either nest the design of estimators against a backdrop of DDC, or vice versa. More specifically, the DDC construct focuses on the metric and the threshold that pick data of significance in terms of better estimation performance guarantees; in diametrical contrast, the design of an estimator is primarily concerned with the best performance criterion that accommodates the characteristics and effects of distinct DDC schemes. Another line of research deals with the joint decision and estimation (JDE) problem [23] to optimally calibrate the tradeoff between estimation quality and communication expenses. In the meantime, the community also strive to address the problems arising from multi-sensor settings and nonlinear system/non-Gaussian noises, as well as network-induced issues (quantization, packet loss, delay, channel fading and cyber attacks, etc.). The stability assessments of estimators and fusers that ensure the applicability in practical control systems have also been studied.

The scope of this dissertation is limited to the discrete-time linear systems. The relevant work can be broadly sorted into the following domains.

Common DDC strategies

A cluster of DDC strategies are based on the idea of existing event-based sampling schemes including "Send-on-Delta" (SOD) [24] and "Integral sampling" [25]. SOD, also known as deadbands [26], was proposed in [18, 24, 27]. With SOD, transmission is triggered only when the differences between the current measurement and the last transmitted one is greater than a given threshold. It serves as a prototype model and has made its way into a range of state estimation problems [28–30]. Based on the idea of "integral sampling", [31] brought forward the "Send-on-Area" (SOA) approach.

Some DDC strategies design the metric from the statistical point of view. [32] built the metric based on the Kullback-Leibler (KL)-divergence between the updated posterior and its predicted value. This metric was also studied in [33, 34]. [35] proposed a variance-based triggering mechanism to make the communication decisions off-line, its effectiveness being partially affected by lack of real-time information notwithstanding [36].

Another class of DDC strategies constructs the metric based on the innovation. For estimation, innovation carries new information in the current data that is not in old data [37]. A small innovation indicates a small error for the corresponding prediction, and thus is better not transmitted to save communication resources. A large innovation suggests that it carries much information for estimation. [38] and [39] showed that estimators based on quantized measurement innovation exhibit similar performance as that of the Kalman filter. Based on the measurement innovation,

[40] suggested a deterministic thresholding scheme that was developed from SOD. With the measurement/measurement innovation, [41] designed a stochastic thresholding scheme. The stochastic approaches, in spite of the additional uncertainties, have the advantage of preserving Gaussianity of the posterior pdf of the state, leading to an array of applications [42–44]. Estimate innovation serves our goal of filtering better than measurement innovation since it is more direct and natural for estimation. [45] modeled data sending times as jumps of an integer random process. For both measurement innovation and estimate innovation, deterministic policies have been thoroughly investigated [46–50].

In addition to the study of metric, the community is also researching various thresholding policies. These policies can be either static or dynamic. The static threshold can be fixed or follows an expression on the sensor measurements/estimates; they can be sorted into the constant threshold, the instantaneous measurement- or estimate-dependent threshold and the released measurement- or estimate-dependent threshold [51]. The dynamic thresholding policies—by inserting an additional internal dynamic variable—produce larger average inter-event time in comparison with the conventional static schemes [52–57].

Research on data-driven estimation

An important question is how to exploit the information from the non-transmission region, per DDC schemes, to improve estimation performance. An ensemble of hybrid update estimators have been designed to ingest either outright transmitted signal at the times of transmission or general knowledge regarding the non-transmission region in the otherwise case. [34, 58] modeled the knowledge of the region as a non-stochastic set-membership representation. A large chunk of the literature models the region as a set of stochastic measurements, rendering a truncated representation of the distribution. This poses problems in Bayesian updates because the posterior distribution of the system state becomes non-Gaussian even for a linear-Gaussian system. In this case, the exact MMSE estimators are intractable computationally. As such, approximations are employed, among which Gaussian approximation and Gaussian summation approximation are widely adopted [33, 40, 48, 49, 59]. [60] offered a more accurate prescription by introducing the generalized closed skew normal (GCSN) distribution.

In practical applications, model properties are often unavailable or amiss, calling for robust filters that tolerate such uncertainties. H_∞ estimator, also called the minimax estimator, minimizes the worst-case estimation error [8]. Another robust risk-sensitive filtering minimizes the expectation of the exponential of the squared estimation error multiplied by a risk-sensitive parameter [61–63]. Under the framework of DDC, these two robust estimators have been steadily gaining grounds in recent years [55, 64–69]. Besides, the set-membership estimation [70, 71], the moving horizon estimation [72], the estimation problems for HMM [73] and energy harvesting sensors [74, 75] and the problem of estimation and control [76] were also studied.

The term "stability" in the area of state estimation refers to a non-diverging error covariance, which in turn means a non-diverging estimation error. The stability study of estimators and fusers ensures their applicability in practical settings [41, 49, 77–79].

Research on estimation fusion with DDC

Estimation fusion is investigated for both systems with and without fusion centers. Under centralized fusion architectures, SOD, variance-based, stochastic thresholding and deterministic thresholding schemes were studied in [28], [35], [80] and [81], respectively. For distributed fusion constructs, innovation-based, variance-based and information-based triggering schemes were explored in [71, 82], [74, 83] and [84]. It is noteworthy that [77, 85] considered both architectures. For a networked system without a fusion center, KL divergence-based triggering, innovation-based and dynamic thresholding schemes were analyzed in [78, 86], [50, 87, 88] and [54, 70, 89]. The community also looked into stability, robust estimation and set-membership estimation for systems with multiple sensors [71, 77, 78, 80, 90].

Optimal JDE

Another line of research on data-driven estimation treats limited communication resource quantitatively in the joint design, which formulates the problems in the framework of optimal estimation by incorporating these limited resource. Three general routes are pursued to address this JDE problem:

I. Minimizing the communication cost with constraints on estimation error. [91] introduced a distributed greedy heuristic for the global minimization problem such that sensors determine their communication policies using local information available.

II. Minimizing the estimation error constrained by the communication cost. [92–94] studied the optimization problem in a finite time horizon using dynamic programming. [95] gave a sub-optimal solution via generalized geometric programming optimization techniques.

III. Minimizing a weighted sum of estimation error and communications cost. [96] framed the problem as one of an infinite time horizon and solved it using dynamic programming. [97] proposed an iterative algorithm over a finite horizon. [98] considered the data-driven estimation problem with packet drops. [99] studied an optimal event-triggered estimation problem under a stochastic thresholding scheme.

Research on data-driven estimation with network induced issues

Communication, upon being incorporated in the system, inevitably introduces additional complexities, including quantization, delays, data losses, cyber attacks, etc. [67, 83, 100, 101] considered the impact of data loss. Issues of delay were examined in [57, 68, 69, 102, 103]. [104] mulled over measurement quantization and random sensor failures. [84, 105] proposed designs that are able to distinguish between an actual measurement and noise at the remote estimator. Deception attacks were accounted for in system design by [106].

It is noteworthy that efficient innovation-driven schemes that transmit significant data for better performance still await thorough investigation. In the meantime, estimation fusion with DDC and the optimal JDE problem are still in their early phases of research. We strive to understand the following problems in this dissertation:

Problem I The joint construction of a DDC scheme and an optimal estimator/fuser for the one-sensor and the multi-sensor cases.

Problem II The design of an efficient innovation-driven scheme that transmits significant data for better performance.

Problem III The stability analysis of the remote estimator and fuser with DDC.

Problem IV The design of a DDC scheme and an estimator to quantitatively balance system performance and limited communication resources.

1.4 Main contributions and outlines

This dissertation focuses on the systemic design of proper estimation as well as fusion techniques and DDC schemes to infer the state of a discrete-time linear dynamical system over a network. Our work effectively helps extract and share key information over networks to enhance estimates of system states. The main contributions of our research are as follows:

1) Based on one sensor estimation, centralized and distributed estimation fusion architectures, respectively, we jointly construct DDC schemes and optimal estimators/fusers. The proposed DDC schemes take into account multiple network structures and the influence of global information. They utilize the importance of innovation for estimation and can achieve a trade-off between communication costs and estimation performance. The proposed corresponding MMSE estimator/fuser and the MMSE-optimal WLS fuser have closed-form representations and optimally use information in the triggering decisions, with theoretical rigor, to boost estimation performance, even for the case in which each sensor has its own triggering policy.

2) We propose new DDC schemes based on cumulative estimate innovation to smartly pick key information and derive the corresponding MMSE estimator/fuser, for both the single-sensor case and the multi-sensor case. First, the cumulative estimate innovation, a new triggering variable, is introduced to reflect the importance of data. The proposed scheme is path-dependent and has the cumulative property. It can detect oscillation and does not suffer from the problem of "no transmission for a long time". Also, the cumulative estimate innovation is less random than the instant innovation. As such, the threshold is more easily determined for a better solution of this mixed estimation-communication problem. Moreover, what is transmitted contains more information for estimation than otherwise, so the degradation of estimation quality due to reduced communication is minimal. What's more, for the remote center, "no transmission" is informative, and this informa-

tion is used optimally to improve estimation performance. The estimator/fuser are optimal under well justified assumptions and have guaranteed stability.

3) We quantitatively calibrate the tradeoff between estimation quality and communication expenses to obtain an optimal data-driven estimator. Firstly, we propose a hybrid data-time-driven communication scheme and derive the MMSE estimator accordingly. We have proved, with due mathematical rigor, that the estimation error matrix—conditioned on the information carried by the entire "no data transmission" duration, does not exceed the one conditioned only on the information carried by the last "no data transmission" timestamp. We further give its workable tight upper bound. Secondly, to quantify the tradeoff between estimation quality and communication expenses, we define the optimization criterion based on the expected total discounted cost over the infinite horizon. We are able to ascertain a highly nontrivial finding—the stochastic linear system under our communication scheme is representable as an MDP problem. An iterative algorithm is proposed to find the optimal policy and optimal cost. We also extend our results to the optimal average cost problem. Lastly, we prove a degeneracy property that is peculiar to the optimal triggering policy. The computational overhead of the proposed design method is less sensitive to the system dimension compared with that of existing algorithms in the literature.

4) The proposed DDC schemes under our metric have easy setup and can seamlessly integrate different dynamical systems, because the optimal decision policy is algorithmic attainable. The stability study of estimators and fusers ensures their applicability in practical control systems. The designed DDC schemes and estimators/fusers can provide optimal solutions for state estimation over sensor networks; also they are reliable and easy to implement across multiple sensor network architectures.

The remainder of the dissertation, consisting of five chapters, is structured as follows.

In chapter 2, we work primarily on **Problem I**. We study the problem of jointly constructing DDC schemes and optimal estimators/fusers for a discrete-time stochastic linear system based on data collected from multiple sensors with limited communication resources.

Chapter 3 is concerned with **Problem I, II and III**. In this chapter, we propose new DDC schemes based on cumulative estimate innovation and derive the corresponding stable MMSE estimator/fuser.

Chapter 4 deals with **Problem IV**. In this chapter, we consider the design of DDC scheme and derive the corresponding estimator to quantitatively calibrate the tradeoff between estimation quality and communication expenses.

Chapter 5 summarizes our work and discusses future research.

CHAPTER 2

ESTIMATION FUSION WITH INSTANT-INNOVATION-DRIVEN COMMUNICATION

2.1 Introduction

Estimation fusion is used when multiple sensors are deployed. The limited communication resources already poses challenges in single sensor design, and it was aggravated with the inception of multiple sensors. A proper fusion technique working in synergy with a data-driven communication scheme may pick only the most relevant pieces of information to send, thus ensuring high effectiveness while maintaining lower costs. The literature includes an array of work on estimation fusion with DDC, such as information-based, innovation-driven, variance-based and KL divergence-based communication schemes [71, 77, 78, 80–82, 84, 85]. Among them, innovation-driven communication strategies are hailed as promising ones. [80] considered centralized fusion problem with the stochastic thresholding scheme proposed in [41]. Under the Gaussian assumption, [81] proposed sequential fuser with a measurement-innovation-driven scheme. [77] considered both centralized and distributed fusion architectures with a novel communication scheme based on minimizing the volume of the non-transmission region. [82] suggested a modified version of the covariance intersection algorithm with an estimate-innovation-driven scheme. This chapter contributes our share by studying both centralized and distributed estimation fusion with communication driven by instant innovations for multi-sensor case, potentially opening new avenues for a broad area of applications.

In this chapter, we study the data-driven state estimation fusion problem for a discrete-time stochastic linear system based on data collected from multiple sensors with limited communication resources. For the cases of transmitting measurements and local state estimates, centralized and distributed estimation fusion architectures are considered, respectively. Under these two basic fusion architectures, we co-design DDC schemes based on a normalized innovation vector and corresponding fusion rules in the (approximate) MMSE sense. A simulation example is provided to confirm the effectiveness of the proposed strategies. This chapter has the following main contributions.

1) In the case of transmitting local estimates, we propose a DDC scheme that decides to send or not based on a normalized local estimation innovation vector. The fusion rule is based on the optimal WLS fusion.

2) In the case of transmitting measurements, we present a DDC scheme, which extends the deterministic thresholding scheme of [40] to the multi-sensor case where the system with any single sensor is not necessarily observable. The fusion rule is based on an approximate MMSE estimator.

3) These DDC schemes can achieve a trade-off between communication costs and estimation performance. A feedback loop from the fusion center offers immediately an aerial view about the accuracy of global estimates, thereby helping in the delivery of significant data. These fusion rules can improve the estimation performance by knowing the fact that no transmission of data indicates a small innovation. We also give the relationship between the average communication rate and the pre-defined threshold for the two cases.

The remainder of the chapter is organized as follows. Section 2.2 formulates the multi-sensor state estimation problem and proposes the idea of DDC. Section 2.3 designs a DDC scheme and an approximate MMSE estimator for the case of raw measurement transmission. Section 2.4 designs a DDC scheme and an optimal WLS fuser for the case of local state estimate transmission. Section 2.5 consists of a simulation. A conclusion at the end summarizes the work.

2.2 Problem formulation

Consider a discrete-time stochastic linear system

$$x_{k+1} = Ax_k + w_k, \quad k \in \mathbb{Z}, \tag{2.1}$$

where $x_k \in \mathbb{R}^n$ is the system state, $A \in \mathbb{R}^{n \times n}$ ($n \in \mathbb{N}$) is the system matrix, the initial state is $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$, ansd the driving noise is $w_k \sim \mathcal{N}(0, Q)$.

To remotely estimate the system state, $s \in \mathbb{N}$ sensors are used to measure x_k in Eq. (2.1). Each sensor i ($i \in \{1, 2, \cdots, s\}$) provides a noisy measurement

$$y_k^i = C^i x_k + v_k^i, \tag{2.2}$$

where $y_k^i \in \mathbb{R}^{m_i}$ is the measurement, $C^i \in \mathbb{R}^{m_i \times n}$ ($m_i \in \mathbb{N}$) is the output matrix, and the measurement noise is $v_k^i \sim \mathcal{N}(0, R^i)$. Define $y_k = [(y_k^1)', (y_k^2)', \cdots, (y_k^s)']'$, $C = [(C^1)', (C^2)', \cdots, (C^s)']'$.

Assumption 2.1 *The initial state x_0, the process noise sequence $\langle w_k \rangle$, and the measurement noise sequences $\{\langle v_k^i \rangle, i \in \{1, 2, \cdots, s\}\}$ are independent for each sensor and pairwise uncorrelated across the sensors.*

Assumption 2.2 *(A, C) is observable, but each pair (A, C^i) is not necessarily observable.*

These remote sensors can process their measurements to find local state estimates $\hat{x}_k^1, ..., \hat{x}_k^s$. The raw measurements or local estimates of each sensor i are coded into data packets and transmitted to the fusion center via multiple independent communication channels. The center fuses the data received to estimate the system state. The problem is to design controlled communication schemes for deciding which data should be transmitted and the corresponding fusion rules for each type of data so as to reduce the communication rate, while trying to have good estimation quality.

Since innovation is useful information for estimation, we consider the DDC based on the innovation, where there exists a unique decision variable for each of s sensors to transmit to the fusion center. To be more specific, only important data with its normalized innovation larger than a pre-defined threshold will be transmitted to the fusion center — the rest will not be transmitted, to save communication resources.

We specify γ_k^i as a binary decision variable for sensor i, where $\gamma_k^i = 1$ indicates that sensor i transmits data at time k and $\gamma_k^i = 0$ otherwise. The average communication rate $\alpha^i \in (0, 1)$ from each sensor i to the fusion center is defined as

$$\lim_{t\to\infty}\frac{1}{t}\sum_{k=1}^{t}\mathbb{E}[\gamma_k^i]=\alpha^i. \tag{2.3}$$

The DDC scheme is defined as

$$\gamma_k^i=\begin{cases}1 & \text{if } ||\epsilon_k^i||_\infty>\delta^i\\ 0 & \text{otherwise,}\end{cases} \tag{2.4}$$

where ϵ_k^i is the normalized innovation vector for transmitting data of sensor i at time k, and $\delta^i<\infty$ is a pre-defined threshold. “No data transmission” corresponds to $||\epsilon_k^i||_\infty\leq\delta^i$, which can be used to improve estimation performance.

We assume at each time k, q_k sensors with index $j_1,\cdots,j_{q_k}$ do not transmit their data and $s-q_k$ sensors with index $j_{q_{k+1}},\cdots,j_s$ transmit their data. For an arbitrary vector χ, we define $\chi^\Lambda=[(\chi^{j_1})',\cdots,(\chi^{j_{q_k}})']'$ and $\chi^\Gamma=[(\chi^{j_{q_{k+1}}})',\cdots,(\chi^{j_s})']'$, where $j_1,\cdots,j_s\in\{1,2,\cdots,s\}$. We also assume that the fusion center has the information about the system parameters.

2.3 Centralized Fusion with measurement transimisson

2.3.1 Data-driven communication scheme

In the measurement transmission case, collecting our decision variables over s sensors, we have $\gamma_k=\text{diag}(\gamma_k^1 I_{m_1},\gamma_k^2 I_{m_2},\cdots,\gamma_k^s I_{m_s})$. Then the available data at the fusion center at each time k is $\mathcal{I}_k=\{\gamma_0 y_0,\gamma_1 y_1,\cdots,\gamma_k y_k\}\bigcup\{\gamma_0,\gamma_1,\cdots,\gamma_k\}$.

Given the available data, the MMSE estimator at the fusion center is

$$\hat{x}_{k|k-1}=\mathbb{E}[x_k|\mathcal{I}_{k-1}], \tag{2.5}$$

$$P_{k|k-1}=\mathbb{E}[(x_k-\hat{x}_{k|k-1})(x_k-\hat{x}_{k|k-1})'|\mathcal{I}_{k-1}], \tag{2.6}$$

$$\hat{x}_k=\mathbb{E}[x_k|\mathcal{I}_k], \tag{2.7}$$

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)'|\mathcal{I}_k], \tag{2.8}$$

where $\hat{x}_{k|k-1}$ and $\hat{x}_k$ are the prior and posterior MMSE estimates, respectively.

The innovation of sensor i is z_k^i, where $z_k^i = y_k^i - \mathbb{E}[y_k^i|\mathcal{I}_{k-1}] = y_k^i - C^i\hat{x}_{k|k-1}$. Given $\mathcal{I}_{k-1}$, we have $z_k^i \sim \mathcal{N}(0, S_k^i)$, where $S_k^i = C^i P_{k|k-1}(C^i)' + R^i$.

Since $S_k^i > 0$, there exists a unitary matrix $U_k^i \in \mathbb{R}^{m_i \times m_i}$ such that

$$(U_k^i)' S_k^i U_k^i = \Lambda_k^i, \tag{2.9}$$

where $\Lambda_k^i = \text{diag}(\lambda_1^i, \lambda_2^i, \cdots, \lambda_{m_i}^i) \in \mathbb{R}^{m_i \times m_i}$ and $\lambda_1^i, \lambda_2^i, \cdots, \lambda_{m_i}^i$ are the eigenvalues of S_k^i. Define $F_k^i \in \mathbb{R}^{m_i \times m_i}$ as $F_k^i = (\Lambda_k^i)^{-\frac{1}{2}}(U_k^i)'$. Obtain the normalized innovation vector ϵ_k^i as

$$\epsilon_k^i = F_k^i z_k^i. \tag{2.10}$$

A DDC scheme for the measurement transmission case is designed as follows:

1) At each time k, the fusion center calculates the predicted measurement $\hat{y}_{k|k-1}^i$ and the transformation matrix F_k^i, and sends them to sensor i, where $\hat{y}_{k|k-1}^i = \mathbb{E}[y_k^i|\mathcal{I}_{k-1}] = C^i\hat{x}_{k|k-1}$.

2) Each sensor i calculates ϵ_k^i of Eq. (2.10) and generates γ_k^i according to the indicator function (2.4).

The design architecture is given in Fig. 2.1.

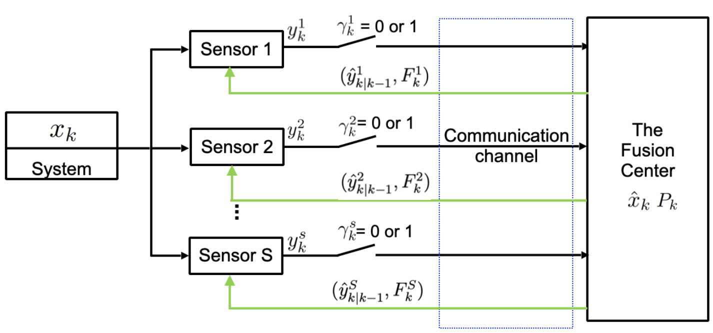


Figure 2.1: The system architecture for centralized fusion with measurement transmission.

Remark 2.1 *$\hat{y}^i_{k|k-1}$ and F^i_k are calculated at the fusion center since system parameters of other sensors and $\mathcal{I}_{k-1}$ are not available for sensor i. To save energy of each sensor, F^i_k is still calculated by the fusion center even if sensor i has the system parameters of other sensors.*

2.3.2 Approximate MMSE estimator

When $\gamma^i_k = 0$, estimation at the fusion center becomes a nonlinear filtering problem, as information about the system is typically a nonlinear function of measurements, i.e., $||\epsilon^i_k||_\infty \leq \delta^i < \infty$. It is known that the exact MMSE estimator under a nonlinear measurement function is computationally intractable in general [107].

Before giving an approximate MMSE estimator, we state a few assumptions, definitions and preliminary results. To derive a recursive filter, we have two assumptions:

Assumption 2.3 $f_{x_k}(x_k|\mathcal{I}_{k-1}) \sim \mathcal{N}(\hat{x}_{k|k-1}, P_{k|k-1})$.

Assumption 2.4 *ϵ^i_k are mutually uncorrelated across different sensors at any time k.*

Let the set $\Omega = \{\epsilon^i_k : ||\epsilon^i_k||_\infty \leq \delta^i, i \in \{j_1, \cdots, j_{q_k}\}\}$ represent the information about the data of sensor i with $\gamma^i_k = 0$. Then $\mathcal{I}_k$ can be re-denoted as $\mathcal{I}_k = \mathcal{I}_{k-1} \cup y^\Gamma_k \cup \Omega$, i.e., the union of the information at time $k-1$, the received measurements at time k, and the information about data not transmitted at time k. Re-indexing the output matrix C and the covariance R of the measurement noise, we have $C_k = [(C^{j_1})', \cdots, (C^{j_{q_k}})', (C^{j_{q_k+1}})', \cdots, (C^{j_s})']'$ and $R_k = \text{diag}(R^{j_1}, \cdots, R^{j_{q_k}}, R^{j_{q_k+1}}, \cdots, R^{j_s})$. Define

$$\Psi_k = \text{diag}(r^{j_1}_k I_{m_{j_1}}, \cdots, r^{j_s}_k I_{m_{j_s}}), \tag{2.11}$$

$$\Gamma_k = \{\Psi_k\}_0, \tag{2.12}$$

$$\Lambda_k = \{I_m - \Psi_k\}_0, \tag{2.13}$$

where $m = \sum_{i=1}^s m_i$.

Lemma 2.1 *Under Assumptions 2.1 and 2.3, we have $f(x_k|\mathcal{I}_{k-1}, z_k^\Lambda, y_k^\Gamma) \sim \mathcal{N}(\mu_1, \sigma_1^2)$, where*

$$\mu_1 = \hat{x}_{k|k-1} + L_k^\Gamma(y_k^\Gamma - \Gamma_k C_k \hat{x}_{k|k-1}) + L_k^\Lambda z_k^\Lambda, \tag{2.14}$$

$$\sigma_1^2 = P_{k|k-1} - P_{k|k-1} C_k'(S_k)^{-1} C_k P_{k|k-1}, \tag{2.15}$$

$$L_k^\Gamma = P_{k|k-1} C_k'(S_k)^{-1} \Gamma_k', \tag{2.16}$$

$$L_k^\Lambda = P_{k|k-1} C_k'(S_k)^{-1} \Lambda_k, \tag{2.17}$$

$$S_k = C_k P_{k|k-1} C_k' + R_k. \tag{2.18}$$

Proof: Under Assumption 2.1, $f([x_k', z_k^{\Lambda'}, y_k^{\Gamma'}]')$ is jointly Gaussian distribution given $\mathcal{I}_{k-1}$ and $f([x_k', z_k^{\Lambda'}, y_k^{\Gamma'}]'|\mathcal{I}_{k-1}) \sim \mathcal{N}(\mu, \sigma^2)$, where

$$\mu = \begin{bmatrix} \hat{x}_{k|k-1} \\ \mathbf{0} \\ \Gamma_k C_k \hat{x}_{k|k-1} \end{bmatrix}, \tag{2.19}$$

$$\sigma^2 = \begin{bmatrix} P_{k|k-1} & P_{k|k-1} C_k' \Lambda_k' & P_{k|k-1} C_k' \Gamma_k' \\ \Lambda_k C_k P_{k|k-1} & \Lambda_k S_k \Lambda_k' & \Lambda_k S_k \Gamma_k' \\ \Gamma_k C_k P_{k|k-1} & \Gamma_k S_k \Lambda_k' & \Gamma_k S_k \Gamma_k' \end{bmatrix}, \tag{2.20}$$

According to Example 3.2 of [37], given a jointly Gaussian distribution $f([x_k', z_k^{\Lambda'}, y_k^{\Gamma'}]'|\mathcal{I}_{k-1})$, it is easy to obtain distribution of $f(x_k|\mathcal{I}_{k-1}, y_k^\Gamma, z_k^\Lambda)$, which is also a Gaussian distribution show in Lemma 2.1.

Remark 2.2 *In Lemma 2.1, we consider that $\{j_1, \cdots, j_{q_k}\}$ and $\{j_{q_{k+1}}, \cdots, j_s\}$ form an arbitrary partition of the set $\{1, \cdots, s\}$: $\{j_1, \cdots, j_{q_k}\} \cap \{j_{q_k} + 1, \cdots, j_s\} = \emptyset$ and $\{j_1, \cdots, j_{q_k}\} \cup \{j_{q_{k+1}}, \cdots, j_s\} = \{1, \cdots, s\}$.*

Lemma 2.1 gives the distribution of x_k under the information at time $k-1$, the measurements of the sensors indexed by $j_{q_{k+1}}, \cdots, j_s$ and the innovation of the sensors indexed by $j_1, \cdots, j_{q_k}$. The effects of the two arbitrary disjoint and collectively exhaustive groups of data on x_k are uncoupled.

Lemma 2.2 *Under Assumptions 2.1, 2.3 and 2.4,*

$$\mathbb{E}[\epsilon_k^{\Lambda}|\mathcal{I}_{k-1}, y_k^{\Gamma}, \Omega] = \mathbf{0}, \tag{2.21}$$

$$\mathbb{E}[\epsilon_k^{\Lambda}\epsilon_k^{\Lambda'}|\mathcal{I}_{k-1}, y_k^{\Gamma}, \Omega] = \textit{diag}([1-\beta(\delta^{j_1})]I_{m_{j_1}}, \cdots, [1-\beta(\delta^{j_{q_k}})]I_{m_{j_{q_k}}}), \tag{2.22}$$

where $\beta(\delta) = \frac{2}{\sqrt{2\pi}}\delta\exp(-\frac{\delta^2}{2})(1-2Q(\delta))^{-1}$, $Q(\delta) = \int_{\delta}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})dx$.

Proof: Given $\mathcal{I}_{k-1}$, $\epsilon_k^i = F_k^i z_k^i \sim \mathcal{N}(0, I_{m_i})$. Let $\epsilon_k^{i,l}$ be the lth element of ϵ_k^i. Then, for $i, j \in \{1, \cdots, q_k\}$,

$$\begin{aligned}\mathbb{E}[\epsilon_k^{i,l}\epsilon_k^{i,p}|\mathcal{I}_{k-1}, y_k^{\Gamma}, \Omega] &= \mathbb{E}[\epsilon_k^{i,l}(\epsilon_k^{i,p})'|\mathcal{I}_{k-1}, y_k^{\Gamma}, |\epsilon_k^{i,l}| \le \delta^i, |\epsilon_k^{i,p}| \le \delta^i] \\ &= 0 \qquad l \neq p, \end{aligned}\tag{2.23}$$

$$\mathbb{E}[(\epsilon_k^{i,l})^2|\mathcal{I}_{k-1}, y_k^{\Gamma}, \Omega] = \mathbb{E}[(\epsilon_k^{i,l})^2|\mathcal{I}_{k-1}, y_k^{\Gamma}, |\epsilon_k^{i,l}| \le \delta^i] = 1-\beta(\delta^i), \tag{2.24}$$

$$\begin{aligned}\mathbb{E}[\epsilon_k^{i}(\epsilon_k^{j})'|\mathcal{I}_{k-1}, y_k^{\Gamma}, \Omega] &= \mathbb{E}[\epsilon_k^{i}(\epsilon_k^{j})'|\mathcal{I}_{k-1}, y_k^{\Gamma}, ||\epsilon_k^i||_\infty \le \delta^i, ||\epsilon_k^j||_\infty \le \delta^j] \\ &= \mathbf{0} \qquad i \neq j. \end{aligned}\tag{2.25}$$

Lemma 2.2 gives the mean and covariance of the normalized innovation of the non-transmitted measurements. Since only the measurement whose normalized innovation ϵ_k^i is inside a pre-defined region is not transmitted, each entry of the covariance of ϵ_k^i is less than 1 since $1-\beta(\delta^i) < 1$ when $\delta^i < \infty$ and can be much smaller than 1.

Approximations 2.1, 2.3 and 2.4 lead to a simple estimator, given by the following theorem.

Theorem 2.1 *Consider fusion with the DDC scheme for the measurement transmission case. Under Assumptions 2.1, 2.3 and 2.4, an approximate MMSE estimator is given recursively as follows:*

1) Time update:

$$\hat{x}_{k|k-1} = A\hat{x}_k, \tag{2.26}$$

$$P_{k|k-1} = AP_kA' + Q, \tag{2.27}$$

2) Measurement update:

$$\hat{x}_k = \hat{x}_{k|k-1} + L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}), \tag{2.28}$$

$$P_k = P_{k|k-1} - P_{k|k-1}C_k^{'} S_k C_k P_{k|k-1} + L_k^{\Lambda} M_k (L_k^{\Lambda})', \tag{2.29}$$

where $M_k =$ *diag*$([1-\beta(\delta^{j_1})]S_k^{j_1}, \cdots, [1-\beta(\delta^{j_{q_k}})]S_k^{j_{q_k}})$, *and* L_k^{Γ}, y_k^{Γ}, Γ_k, C_k, S_k, L_k^{Λ}, $\beta(\delta^i)$, S_k^i *were defined before.*

Proof: :

$$\begin{aligned}
\hat{x}_k &= \mathbb{E}[x_k|I_k] = \mathbb{E}[x_k|I_{k-1}, y_k^{\Gamma}, \Omega] \\
&= \mathbb{E}[\mathbb{E}[x_k|I_{k-1}, y_k^{\Gamma}, \Omega, z_k^{\Lambda} = (F_k^{\Lambda})^{-1}\epsilon_k^{\Lambda}]|I_{k-1}, y_k^{\Gamma}, \Omega] \\
&= \int_{\Omega} (\hat{x}_{k|k-1} + L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}) + L_k^{\Lambda}(F_k^{\Lambda})^{-1}\epsilon_k^{\Lambda}) f_{\epsilon_k^{\Lambda}}(\epsilon|I_{k-1}, y_k^{\Gamma}, \Omega) d\epsilon \\
&= (\hat{x}_{k|k-1} + L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1})) \int_{\Omega} f_{\epsilon_k^{\Lambda}}(\epsilon|I_{k-1}, y_k^{\Gamma}, \Omega) d\epsilon + L_k^{\Lambda}(F_k^{\Lambda})^{-1} \int_{\Omega} \epsilon_k^{\Lambda} f_{\epsilon_k^{\Lambda}}(\epsilon|I_{k-1}, y_k^{\Gamma}, \Omega) d\epsilon \\
&= \hat{x}_{k|k-1} + L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}) + L_k^{\Lambda}(F_k^{\Lambda})^{-1}\mathbb{E}[\epsilon_k^{\Lambda}|I_{k-1}, y_k^{\Gamma}, \Omega] \\
&= \hat{x}_{k|k-1} + L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}),
\end{aligned} \tag{2.30}$$

where $F_k^{\Lambda} = \text{diag}(F_k^1, \cdots, F_k^{q_k})$.

$$\begin{aligned}
P_k &= \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)'|I_k] \\
&= \mathbb{E}[(x_k - \hat{x}_{k|k-1} - L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}))(x_k - \hat{x}_{k|k-1} - L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}))'|I_k] \\
&= \mathbb{E}[((e_{k|k-1} - L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^{\Lambda} z_k^{\Lambda}) + L_k^{\Lambda} z_k^{\Lambda}) \\
&\times \; ((e_{k|k-1} - L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^{\Lambda} z_k^{\Lambda}) + L_k^{\Lambda} z_k^{\Lambda})'|I_k] \\
&= \mathbb{E}[(e_{k|k-1} - L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^{\Lambda} z_k^{\Lambda})(e_{k|k-1} - L_k^{\Gamma}(y_k^{\Gamma} - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^{\Lambda} z_k^{\Lambda})'|I_k]
\end{aligned}$$

$$
\begin{aligned}
&+\quad \mathbb{E}[L_k^\Lambda z_k^\Lambda (L_k^\Lambda z_k^\Lambda)'|I_k] \\
&+\quad \mathbb{E}[(e_{k|k-1} - L_k^\Gamma(y_k^\Gamma - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^\Lambda z_k^\Lambda)(L_k^\Lambda z_k^\Lambda)'|I_k] \\
&+\quad \mathbb{E}[(e_{k|k-1} - L_k^\Gamma(y_k^\Gamma - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^\Lambda z_k^\Lambda)'(L_k^\Lambda z_k^\Lambda)|I_k], \qquad (2.31)
\end{aligned}
$$

where $e_{k|k-1} = x_k - \hat{x}_{k|k-1}$.

Using Lemma 2.1, we have

$$
\begin{aligned}
&\mathbb{E}[(e_{k|k-1} - L_k^\Gamma(y_k^\Gamma - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^\Lambda z_k^\Lambda)(e_{k|k-1} - L_k^\Gamma(y_k^\Gamma - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^\Lambda z_k^\Lambda)'|I_k] \\
&= (P_{k|k-1} - P_{k|k-1}C_k'(C_k P_{k|k-1} C_k' + R_k) C_k P_{k|k-1}) \int_\Omega f_{\epsilon_k^\Lambda}(\epsilon | I_{k-1}, y_k^\Gamma, \Omega) d\epsilon \\
&= P_{k|k-1} - P_{k|k-1}C_k'(C_k P_{k|k-1} C_k' + R_k) C_k P_{k|k-1}. \qquad (2.32)
\end{aligned}
$$

Using Lemma 2.2, we have

$$
\begin{aligned}
&\mathbb{E}[L_k^\Lambda z_k^\Lambda (L_k^\Lambda z_k^\Lambda)'|I_k] \\
&= L_k^\Lambda (F_k^\Lambda)^{-1} \mathbb{E}[\epsilon_k^\Lambda \epsilon_k^\Lambda | I_{k-1}, y_k^\Gamma, \Omega] (F_k^\Lambda)^{-'} (L_k^\Lambda)' \\
&= L_k^\Lambda \text{diag}((1-\beta(\delta^{j_1}))(F_k^{j_1})^{-1}(F_k^{j_1})^{-'}, \cdots, (1-\beta(\delta^{j_{q_k}}))(F_k^{j_{q_k}})^{-1}(F_k^{j_{q_k}})^{-'})(L_k^\Lambda)' \\
&= L_k^\Lambda M_k L_k^{\Lambda'}. \qquad (2.33)
\end{aligned}
$$

Using Lemma 2.1, we have

$$
\begin{aligned}
&\mathbb{E}[(e_{k|k-1} - L_k^\Gamma(y_k^\Gamma - \Gamma_k C_k \hat{x}_{k|k-1}) - L_k^\Lambda z_k^\Lambda)(L_k^\Lambda z_k^\Lambda)'|I_k] \\
&= \mathbb{E}[\mathbb{E}[(e_{k|k-1} - L_k^\Gamma(y_k^\Gamma - \Gamma_k C_k \hat{x}_{k|k-1}))(L_k^\Lambda z_k^\Lambda)'|I_{k-1}, y_k^\Gamma, \Omega, z_k^\Lambda = (F_k^\Lambda)^{-1}\epsilon_k^\Lambda]|I_{k-1}, y_k^\Gamma, \Omega] \\
&- \mathbb{E}[L_k^\Lambda z_k^\Lambda (L_k^\Lambda z_k^\Lambda)'|I_k] \qquad (2.34) \\
&= \mathbb{E}[L_k^\Lambda z_k^\Lambda (L_k^\Lambda z_k^\Lambda)'|I_{k-1}, y_k^\Gamma, \Omega] - \mathbb{E}[L_k^\Lambda z_k^\Lambda (L_k^\Lambda z_k^\Lambda)'|I_k] \\
&= 0.
\end{aligned}
$$

Comparing with the standard LMMSE estimator (the Kalman filter), which is based on the measurements of all sensors, the difference lies in the measurement update. The posterior estimate $\hat{x}_k$ is recursive based only on the received measurements y_k^{Γ}. Its error covariance P_k is updated in the same form as in the standard Kalman filter but with an additional error covariance $L_k^{\Lambda} M_k (L_k^{\Lambda})'$ for the non-transmitted measurements.

This additional error covariance arises due to the uncertainty in the non-transmitted measurements. According to the communication scheme, only the measurement whose normalized innovation ϵ_k^i is inside a pre-defined region will not be transmitted. Therefore, at the fusion center, the fact that there is no transmitted measurements is a piece of useful information that indicates a small innovation. A small innovation suggests that the predicted measurement at the fusion center has a small error. In this sense, each non-transmitted measurement is equivalent to sending the prediction along with a reduced error covariance. This leads to the reduction of the uncertainty in state estimation. Thus, P_k is smaller than the estimation error covariance that uses only the transmitted measurements, since the estimator uses in some sense the information in the non-transmitted measurements.

The pre-defined region is determined by the threshold δ^i. $1 - \beta(\delta^i)$ decreases as δ^i decreases, resulting in the reduction of M_k. In other words, the uncertainty in the non-transmitted data at the fusion center decreases. This results in a smaller P_k.

Proposition 2.1 *In the case of measurement transmission, the communication rate α^i is given by*

$$\alpha^i = 1 - [1 - 2Q(\delta^i)]^{m_i}. \tag{2.35}$$

Given a communication rate α^i, the threshold for the DDC scheme in the case of measurement transmission is

$$\delta^i = Q^{-1}\left(\frac{1 - (1 - \alpha^i)^{1/m_i}}{2}\right). \tag{2.36}$$

Proof:

$$
\begin{aligned}
Pr(\gamma_k^i = 0|\mathcal{I}_{k-1}) &= Pr(||\epsilon_k^i||_\infty \leq \delta^i) \\
&= \prod_{j=1}^{m_i} Pr(|\epsilon_k^{i,j}| \leq \delta^i|\mathcal{I}_{k-1}) \\
&= [1 - 2Q(\delta^i)]^{m_i}.
\end{aligned} \tag{2.37}
$$

Therefore, $\alpha^i = Pr(\gamma_k^i = 1|\mathcal{I}_{k-1}) = 1 - [1 - 2Q(\delta^i)]^{m_i}$.

Remark 2.3 *Proposition 2.1 shows that the communication rate* α^i *grows as* δ^i *decreases. A higher communication rate means more energy consumption of sensors. In the meantime, according to Shannon's theorem, the maximum rate at which data can be transmitted over a communication channel (channel capacity) increases with increasing of the bandwidth and the signal-to-noise ratio of the channel. Thus a higher communication rate is more costly.*

2.4 Distributed fusion with local estimate transmission

2.4.1 Data-driven communication scheme

According to Assumption 2.2, system (2.1)-(2.2) is not necessarily observable. In order to avoid a possible divergence in estimation error, it is convenient to focus on the observable subsystem by separating it from the unobservable part (see also [108], [77]). By introducing a suitable nonsingular transformation $(T^i)^{-1}$, the transformed system matrix take the form

$$
\begin{aligned}
&\begin{bmatrix} A_o^i & 0 \\ A_{21}^i & A_{\bar{o}}^i \end{bmatrix} = (T^i)^{-1}AT^i, \quad \begin{bmatrix} B_o^i \\ B_{\bar{o}}^i \end{bmatrix} = (T^i)^{-1}, \\
&\begin{bmatrix} C_o^i & 0 \end{bmatrix} = C^iT^i.
\end{aligned}
$$

The observable part of system (1)-(2) is an n_i-dimensional subsystem with the state $x_k^i = B_o^i x_k$:

$$x_{k+1}^i = A_o^i x_k^i + B_o^i w_k, \tag{2.38}$$

$$y_k^i = C_o^i x_k^i + v_k^i. \tag{2.39}$$

It is well known that for a linear system in the Gaussian noise, the Kalman filtering is optimal in the MMSE sense. Consequently, for each observable subsystem (2.38)-(2.39), the local estimation by the Kalman filter is

$$(\hat{x}_{k|k-1}^i, P_{k|k-1}^i, \hat{x}_k^i, P_k^i, K_k^i) = \mathbf{KF}(A_o^i, C_o^i, B_o^i Q (B_o^i)', R^i, y_k^i). \tag{2.40}$$

with the initial $P_{0|0}^i = B_o^i P_0 (B_o^i)'$ and $\hat{x}_{0|0}^i = B_o^i \bar{x}_0$, where $\mathbf{KF}$ is defined in (1.21) and $S_{o,k}^i = C_o^i P_{k|k-1}^i (C_o^i)' + R^i$.

Define $\hat{x}_k^L = [(\hat{x}_k^1)', (\hat{x}_k^2)', \cdots, (\hat{x}_k^s)']'$. Collecting our decision variables over s sensors, we have $\gamma_k = \text{diag}(\gamma_k^1 I_{n_1}, \gamma_k^2 I_{n_2}, \cdots, \gamma_k^s I_{n_s})$. Then the available data at the fusion center at each time k is $\mathcal{I}_k = \{\gamma_0 \hat{x}_0^L, \gamma_1 \hat{x}_1^L, \cdots, \gamma_k \hat{x}_k^L\} \bigcup \{\gamma_0, \gamma_1, \cdots, \gamma_k\}$.

Given the available data $\mathcal{I}_k$, we define an MMSE fuser at the fusion center as

$$\hat{x}_k = \mathbb{E}[x_k | \mathcal{I}_k], \tag{2.41}$$

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)' | \mathcal{I}_k], \tag{2.42}$$

where $\hat{x}_k$ and P_k are the state estimate and its error covariance matrix, respectively.

The innovation of sensor i is the local estimation innovation $z_k^i = \mathbb{E}[x_k^i | \{y^i\}_0^k] - \mathbb{E}[x_k^i | \mathcal{I}_{k-1}] = \hat{x}_k^i - B_o^i A \hat{x}_{k-1}$.

Since $B_o^i S_k^L (B_o^i)' + P_k^i > 0$, there exists a unitary matrix $U_k^i \in \mathbb{R}^{n_i \times n_i}$ such that

$$(U_k^i)' (B_o^i S_k^L (B_o^i)' + P_k^i) U_k^i = \Lambda_k^i. \tag{2.43}$$

where $S_k^L = AP_{k-1}A' + Q$, $\Lambda_k^i = \text{diag}(\lambda_1^i, \lambda_2^i, \cdots, \lambda_{n_i}^i) \in \mathbb{R}^{n_i \times n_i}$ and $\lambda_1^i, \lambda_2^i, \cdots, \lambda_{n_i}^i$ are the eigenvalues of $B_o^i S_k^L (B_o^i)' + P_k^i$. Define $F_k^i \in \mathbb{R}^{n_i \times n_i}$ as $F_k^i = (\Lambda_k^i)^{-\frac{1}{2}}(U_k^i)'$. Obviously, $(F_k^i)'F_k^i = (B_o^i S_k^L (B_o^i)' + P_k^i)^{-1}$.

Obtain the normalized innovation vector ϵ_k^i as

$$\epsilon_k^i = F_k^i z_k^i. \tag{2.44}$$

A DDC scheme for the local estimate transmission case is designed as follows:

1) At each time k, the fusion center calculates the predicted local estimates $B_o^i A\hat{x}_{k-1}$ and the transformation matrix F_k^i and sends them to sensor i.

2) Each sensor i calculates ϵ_k^i and compares it with a given threshold to decide whether to transmit or not.

Fig. 2.2 gives an illustration of the system design.

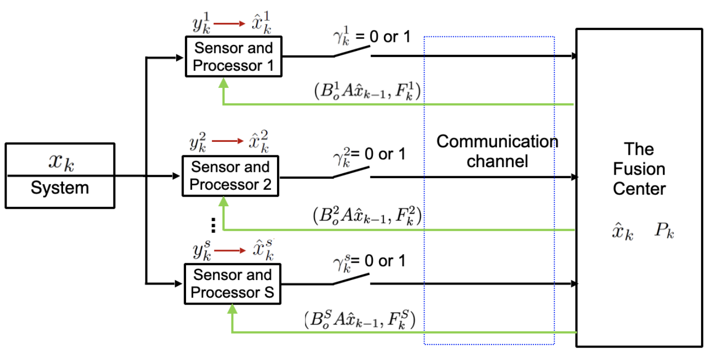


Figure 2.2: The system architecture for distributed fusion with local estimate transmission.

Given the same definition for Ω, $\mathcal{I}_k$ can be re-denoted as $\mathcal{I}_k = \mathcal{I}_{k-1} \cup \hat{x}_k^\Gamma \cup \Omega$, i.e., the union of the information at time $k-1$, the received local estimates at k, and the information about the local estimates not transmitted at time k.

2.4.2 The optimal WLS fuser

Lemma 2.3 *Under Assumption 2.1,*

$$\mathbb{E}[\epsilon_k^i(\epsilon_k^i)' | ||\epsilon_k^i||_\infty \leq \delta^i] = [1-\beta(\delta^i)]I_{n_i}, \tag{2.45}$$

$$\mathbb{E}[z_k^i(z_k^i)' | ||\epsilon_k^i||_\infty \leq \delta^i] = [1-\beta(\delta^i)](B_o^i S_k^L (B_o^i)' + P_k^i). \tag{2.46}$$

Proof:

$$\begin{aligned}
\mathbb{E}[z_k^i] &= \mathbb{E}[\mathbb{E}[x_k^i|\{y^i\}_0^k] - \mathbb{E}[x_k^i|\mathcal{I}_{k-1}]] \\
&= \mathbb{E}[x_k^i] - \mathbb{E}[x_k^i] \\
&= \mathbf{0}, \qquad (2.47) \\
\mathbb{E}[z_k^i(z_k^i)'] &= \mathbb{E}[((x_k^i - \mathbb{E}[x_k^i|\mathcal{I}_{k-1}]) - (x_k^i - \mathbb{E}[x_k^i|\{y^i\}_0^k])) \\
&\quad \times ((x_k^i - \mathbb{E}[x_k^i|\mathcal{I}_{k-1}]) - (x_k^i - \mathbb{E}[x_k^i|\{y^i\}_0^k]))'] \\
&= B_o^i S_k^L (B_o^i)' + P_k^i. \qquad (2.48)
\end{aligned}$$

Thus, $\epsilon_k^i = F_k^i z_k^i \sim \mathcal{N}(0, I_{n_i})$, we have

$$\begin{aligned}
&\mathbb{E}[\epsilon_k^{i,l}(\epsilon_k^{i,p})' | ||\epsilon_k^i||_\infty \leq \delta^i] \\
&= \mathbb{E}[\epsilon_k^{i,l}(\epsilon_k^{i,p})' | |\epsilon_k^{i,l}| \leq \delta^i, |\epsilon_k^{i,p}| \leq \delta^i] \\
&= 0 \qquad l \neq p, \qquad (2.49) \\
&\mathbb{E}[(\epsilon_k^{i,l})^2 | ||\epsilon_k^i||_\infty \leq \delta^i] \\
&= \mathbb{E}[(\epsilon_k^{i,l})^2 | |\epsilon_k^{i,l}| \leq \delta^i] = 1 - \beta(\delta^i). \qquad (2.50)
\end{aligned}$$

Therefore,

$$\mathbb{E}[z_k^i(z_k^i)' | ||\epsilon_k^i||_\infty \leq \delta^i]$$

$$
\begin{aligned}
&= (F_k^i)^{-1}\mathbb{E}[(\epsilon_k^{i,l})^2 | ||\epsilon_k^i||_\infty \le \delta^i](F_k^i)^{-1'} \\
&= (F_k^i)^{-1}(F_k^i)^{-1'}[1-\beta(\delta^i)] \\
&= [1-\beta(\delta^i)](B_o^i S_k^L (B_o^i)' + P_k^i). \qquad (2.51)
\end{aligned}
$$

Theorem 2.2 *Consider fusion with the DDC scheme for the local state estimate transmission case. Under Assumptions 2.1, 2.2 and 2.4, the optimal WLS fuser that minimizes MSE is*

$$
\hat{x}_k = (H'\Sigma_k^{-1}H)^{-1}H'\Sigma_k^{-1}Y_k, \qquad (2.52)
$$

$$
P_k = (H'\Sigma_k^{-1}H)^{-1}. \qquad (2.53)
$$

where $Y_k = col_i(\gamma_k^i\hat{x}_k^i + (1-\gamma_k^i)B_o^iA\hat{x}_{k-1})$, $H = col_i(B_o^i)$, $\Sigma_k = diag(P_k^1 + (1-\gamma_k^1)[1-\beta(\delta^1)](B_o^1S_k^L(B_o^1)' + P_k^1), \cdots, P_k^s + (1-\gamma_k^s)[1-\beta(\delta^s)](B_o^sS_k^L(B_o^s)' + P_k^s))$. *For* s *matrices* $A^1, \cdots, A^s$ *with the same number of columns,* $col_i(A^i) = [(A^1)', \cdots, (A^s)']'$.

Proof: : If $\gamma_k^i = 0$, the available information about sensor i at the fusion center is the predicted local estimate $B_o^iA\hat{x}_{k-1}$ with $||\epsilon_k^i||_\infty \le \delta^i$. We interpret information at the fusion center about each local estimate $\hat{x}_k^i$ as a measurement Y_k^i of the true state x_k collected through the (virtual) communication channel, where

$$
Y_k^i = \begin{cases} B_o^iA\hat{x}_{k-1} = B_o^ix_k + (\hat{x}_k^i - x_k^i) - (\hat{x}_k^i - B_o^iA\hat{x}_{k-1}) & if\ ||\epsilon_k^i||_\infty \le \delta^i \\ \hat{x}_k^i = B_o^ix_k + (\hat{x}_k^i - x_k^i) & otherwise. \end{cases} \qquad (2.54)
$$

Here $(\hat{x}_k^i - x_k^i) - (\hat{x}_k^i - B_o^iA\hat{x}_{k-1})$ and $\hat{x}_k^i - x_k^i$ are considered as the measurement noise u_k^i for the case of transmission and no transmission, respectively. $\hat{x}_k^i - B_o^iA\hat{x}_{k-1}$ is uncorrelated with local estimate error $\hat{x}_k^i - x_k^i$. Summing up all sensors, the available information at fusion center can be treated as originating from the measurement channel

$$
Y_k = H_kx_k + u_k, \qquad (2.55)
$$

where

$$\begin{aligned} Y_k &= col_i(\gamma_k^i \hat{x}_k^i + (1-\gamma_k^i) B_o^i A \hat{x}_{k-1}), &(2.56)\\ H &= col_i(B_o^i), &(2.57)\\ u_k &= col_i(\gamma_k^i(\hat{x}_k^i - x_k^i) + (1-\gamma_k^i)(B_o^i A \hat{x}_{k-1} - x_k^i)). &(2.58)\end{aligned}$$

If $\gamma_k^i = 1$,

$$\begin{aligned} \mathbb{E}[u_k^i | I_k] &= \mathbb{E}[\hat{x}_k^i - x_k^i | I_k] = 0, &(2.59)\\ cov[u_k^i | I_k] &= \mathbb{E}[(\hat{x}_k^i - x_k^i)(\hat{x}_k^i - x_k^i)' | I_k] = P_k^i. &(2.60)\end{aligned}$$

If $\gamma_k^i = 0$, according to the Lemma 2.3 and Assumption 2.4, we have

$$\begin{aligned} \mathbb{E}[u_k^i | I_k] &= \mathbb{E}[(\hat{x}_k^i - x_k^i) - (\hat{x}_k^i - B_o^i A \hat{x}_{k-1}) | I_k], \\ &= 0 &(2.61)\\ cov[u_k^i | I_k] &= \mathbb{E}[((\hat{x}_k^i - x_k^i) - (\hat{x}_k^i - B_o^i A \hat{x}_{k-1}))((\hat{x}_k^i - x_k^i) - (\hat{x}_k^i - B_o^i A \hat{x}_{k-1}))' | I_k] \\ &= \mathbb{E}[(\hat{x}_k^i - x_k^i)(\hat{x}_k^i - x_k^i))' | I_k] + \mathbb{E}[(\hat{x}_k^i - B_o^i A \hat{x}_{k-1})(\hat{x}_k^i - B_o^i A \hat{x}_{k-1})' | I_{k-1}, ||\epsilon_k^i||_\infty \leq \delta^i] \\ &+ \mathbb{E}[(\hat{x}_k^i - x_k^i)(\hat{x}_k^i - B_o^i A \hat{x}_{k-1})' | I_k] + \mathbb{E}[(\hat{x}_k^i - x_k^i)'(\hat{x}_k^i - B_o^i A \hat{x}_{k-1}) | I_k] \\ &= P_k^i + (1 - \beta(\delta^i))(B_o^i (A P_{k-1} A' + Q) B_o^{i'} + P_k^i). &(2.62)\end{aligned}$$

We assume the local estimate error $\hat{x}_k^i - x_k^i$ are mutually uncorrelated for any time k, then $cov[u_k^i u_k^j | I_k] = 0$. Consequently, the optimal WLS fuser that minimizes MSE is

$$\begin{aligned} \hat{x}_k &= (H' \Sigma_k^{-1} H)^{-1} H' \Sigma_k^{-1} Y_k, &(2.63)\\ P_k &= (H' \Sigma_k^{-1} H)^{-1}, &(2.64)\end{aligned}$$

where Σ_k is the covariance of the virtual measurement noise $\Sigma_k = \text{diag}(P_k^i + (1 - \gamma_k^i)(1 -$

$\beta(\delta^i)(B_o^i(AP_{k-1}A' + Q)B_o^{i'}))$.

Y_k in Eq. (2.52) is the available information at the fusion center. It can be treated as a virtual measurement of the true state x_k with a virtual measurement noise u_k, i.e., $Y_k = Hx_k + u_k$, where $u_k = \text{col}_i((\hat{x}_k^i - x_k^i) + (1 - \gamma_k^i)(B_o^i A\hat{x}_{k-1} - \hat{x}_k^i))$. Σ_k in Eqs. (2.52) and (2.53) is the covariance of u_k. Comparing with the general WLS fuser using all sensors' local estimates, the difference lies in the update when $\gamma_k^i = 0$. If $\gamma_k^i = 0$, the available information from sensor i is given by the predicted local estimate $B_o^i A\hat{x}_{k-1}$. The corresponding Σ_k is updated as in the general WLS fuser but with an additional covariance $[1 - \beta(\delta^i)](B_o^i S_k^L (B_o^i)' + P_k^i)$.

This additional covariance arises due to the uncertainty in the non-transmitted data. The communication scheme transmits the local estimates that are far from $B_o^i A\hat{x}_{k-1}$ in the sense that the normalized innovation vector falls outside a pre-defined region. Therefore, at the fusion center, the fact that there is no transmitted local estimates is a piece of useful information that indicates a small innovation. The small innovation suggests that the predicted local estimate at the fusion center has a small error. In this sense, each non-transmitted local estimate is equivalent to sending the prediction with a reduced error covariance. This leads to the reduction of the uncertainty in the non-transmitted local estimates. Thus, Σ_k in Eq. (2.53) is smaller than that which uses only the transmitted local estimates.

The pre-defined region is determined by the threshold δ^i. A smaller δ^i also indicates that the $B_o^i A\hat{x}_{k-1}$ is closer to $\hat{x}_k^i$. Thus error of the predicted local estimates decreases, and this can be seen from the decrease in $1 - \beta(\delta^i)$. The eventual effect is a smaller Σ_k.

Proposition 2.2 *In the case of local estimate transmission, the communication rate α^i is given by*

$$\alpha^i = 1 - [1 - 2Q(\delta^i)]^{n_i}. \tag{2.65}$$

Given a communication rate α^i, the threshold for the DDC scheme of local estimate transmis-

sion is

$$\delta^i = Q^{-1}(\frac{1-(1-\alpha^i)^{1/n_i}}{2}). \tag{2.66}$$

A proof is omitted here since it is similar to that of Proposition 2.1.

Proposition 2.2 shows that the communication rate α^i grows as δ^i decreases.

2.5 Illustrative example

This section gives an example to illustrate the results in this chapter and discusses properties of the communication schemes and fusion strategies.

Consider the system (2.1)-(2.2) with a two-dimensional state and system parameters

$$A = \begin{bmatrix} 0.8 & 0.1 \\ 0 & 0.7 \end{bmatrix}, \qquad Q = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix},$$

$$C^1 = \begin{bmatrix} 1 & 1 \\ 0 & 1.3 \end{bmatrix}, \qquad C^2 = \begin{bmatrix} 0 & 1 \end{bmatrix},$$

$R^1 = 0.8I_2$, $R^2 = 1$, $P_0 = 0.3I_2$, $x_0 = [1, 1]'$.

The following four communication schemes have been compared in the cases of transmitting measurements and local state estimates, respectively.

1) DDC: each sensor i transmits important data with their normalized innovation larger than a pre-defined threshold δ^i. This threshold has been determined by Eq. (2.36) or (2.66) to obtain the desired communication rate α^i.

2) Stochastic communication (SC): each sensor i random transmits data with communication rate α^i.

3) Periodic communication (PC): each sensor i periodically transmits data once every $1/\alpha^i$

time instants.

4) Full communication (FC): each sensor transmits data at each time k, i.e., the communication rate is 1.

For communication schemes DDC, SC and PC, the communication rate α^i is fixed at 0.5 for all sensors in both cases.

2.5.1 Fusion with measurement transmission

Using Proposition 2.1, the pre-defined thresholds of two sensors are 1.05 and 0.67, respectively. By applying Theorem 2.1, the approximate MMSE estimator is obtained and its results are shown in Fig. 2.3, where the true values and their estimates are depicted in the blue and red lines, respectively. It can be seen from Fig. 2.3 that the estimates are close to the true values, and the estimator has good performance.

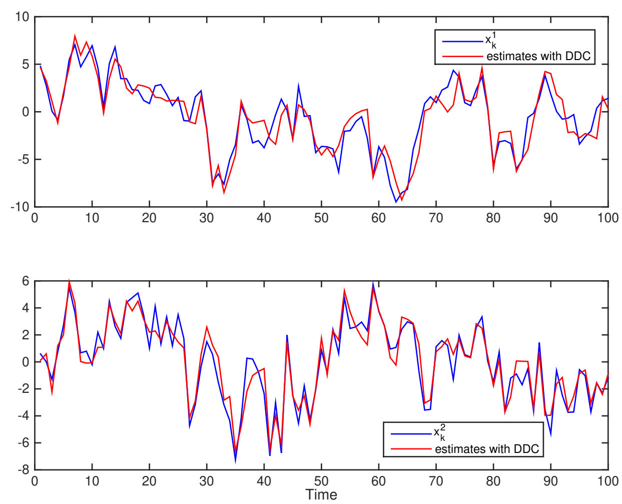


Figure 2.3: The true state components and their estimates in the measurement transmission case

For the four communication schemes, Fig. 2.4 depicts the traces of the estimation error covariance matrices, and Fig. 2.5 depicts the root mean square error (RMSE) over 500 Monte Carlo runs. They show that DDC has errors smaller than SC and PC, and slightly larger than FC. Therefore, in the case of transmitting measurements, our scheme outperforms SC and PC by using the same communication rates. DDC is worse than FC in estimation performance, but it saves communication costs since it requires only half of the communication rate of FC.

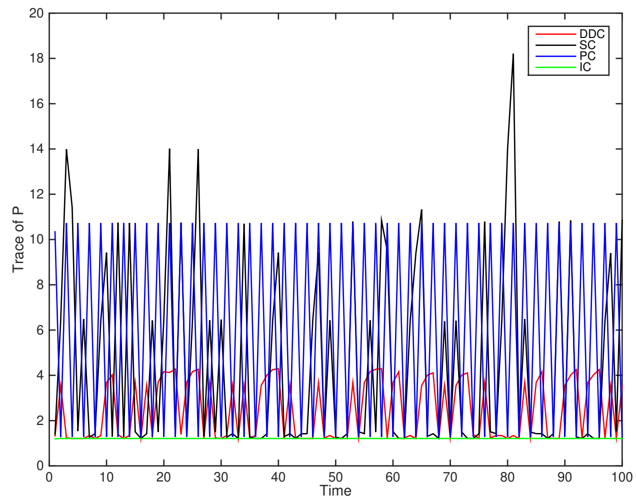


Figure 2.4: Trace of estimation error covariance matrices for the four communication schemes in the measurement transmission case

Table 2.1 gives a detailed comparison for the four communication schemes. First, DDC provides a notable improvement in estimation performance over SC and PC under the same communication rates. With respect to SC, the error reduction is 26.36% for the time average RMSE and 50.15% for the maximum RMSE. With respect to PC, the reduction is 35.46% for the time average RMSE and 37.41% for the maximum RMSE. Second, FC has better estimation performance than DDC. It provides a 32.1% decrease for the time average RMSE and 31.36% decrease for the maximum RMSE. But DDC gives communication cost savings since it reduce the communication rate by half.

Table 2.1: Comparison for the four communication schemes in the measurement transmission case

Communication schemes	**DDC**	**SC**	**PC**	**FC**
time averaged RMSE	1.62	2.20	2.51	1.10
maximum RMSE	1.69	3.39	2.70	1.16

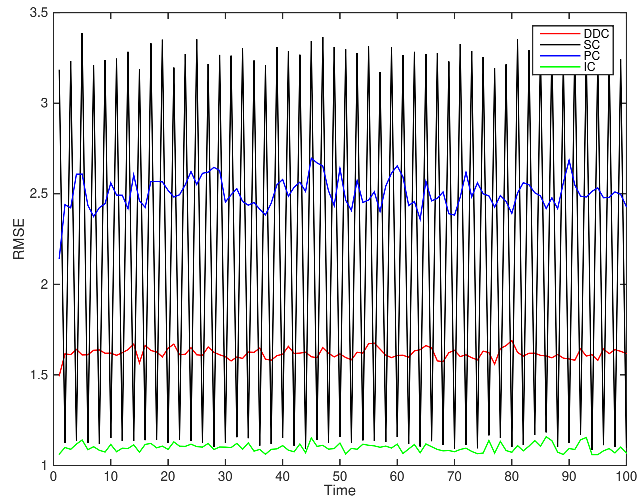


Figure 2.5: RMSE for the four communication schemes in the measurement transmission case

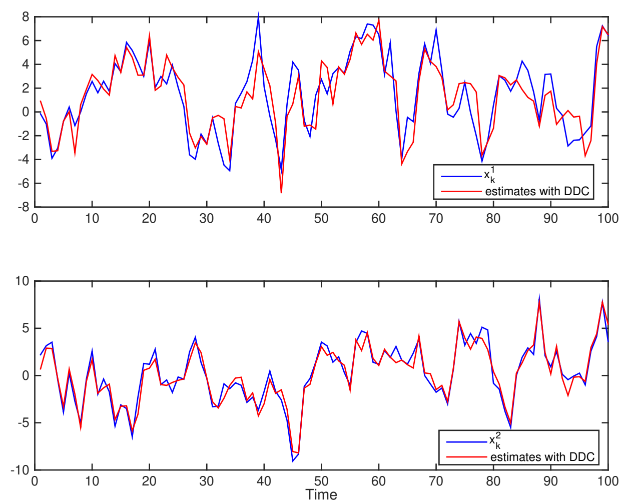


Figure 2.6: The true state components and their estimates in the local estimate transmission case

2.5.2 Fusion with local estimate transmission

Using Proposition 2.2, the pre-defined thresholds of two sensors are 1.05 and 0.67, respectively. By applying Theorem 2.2, the optimal WLS fuser is obtained and its results are shown in Fig. 2.6, where the true values and their estimates are depicted by the blue and red lines, respectively. It can be seen from Fig. 2.6 that the estimates are close to the true values, and the optimal WLS fuser has good estimation performance.

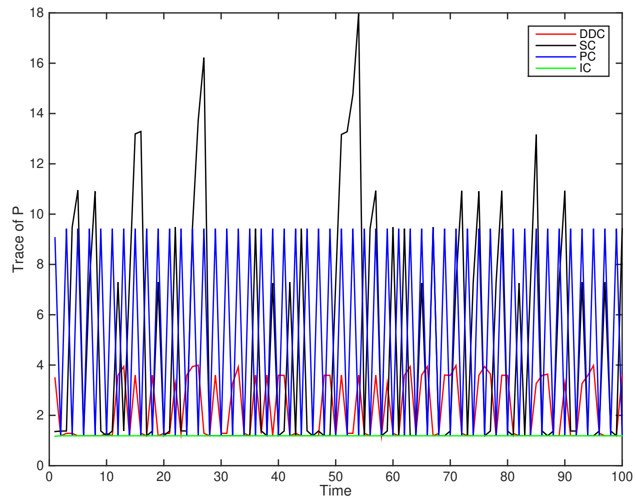


Figure 2.7: Trace of estimation error covariance for the four communication schemes in the local estimate transmission case

For the four communication schemes, Fig. 2.7 depicts the traces of the estimation error covariance matrices, and Fig. 2.8 depicts the RMSE over 500 Monte Carlo runs. They show that, in the case of local estimate transmission, DDC has error smaller than SC and PC, and slightly larger than FC. Therefore, in the case of transmitting local estimates, our scheme outperforms SC and PC by using the same communication rates. DDC is worse than FC in estimation performance, but it saves communication costs since it requires only half of the communication rate of FC.

Table 2.2: Comparison for the four communication schemes in the local estimate transmission case

Communication schemes	**DDC**	**SC**	**PC**	**FC**
time averaged RMSE	1.61	2.20	2.50	1.11
maximum RMSE	1.71	3.52	2.80	1.18

Table 2.2 gives a detailed comparison for the four communication schemes. First, DDC provides a notable improvement in estimation performance over SC and PC under the same communication rates. With respect to SC, the error reduction is 26.82% for the time average RMSE and 51.42% for the maximum RMSE. With respect to PC, the reduction is 35.6% for the time average RMSE and 38.93% for the maximum RMSE. Second, the estimation performance of FC is better than that of DDC. It provides a 31.06% decrease for the time average RMSE and 30.99% decrease

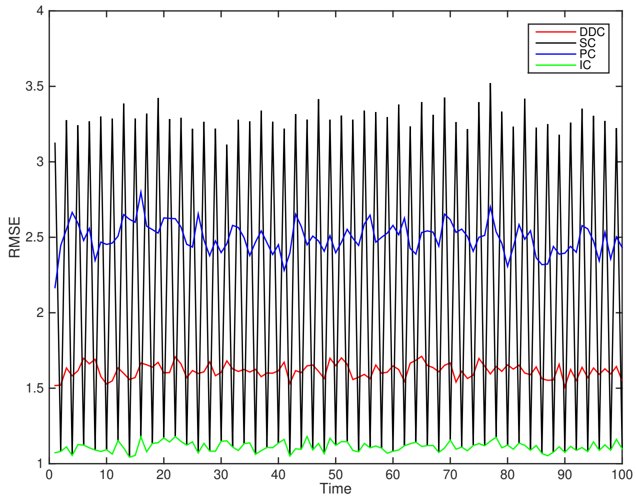


Figure 2.8: RMSE for the four communication schemes in the local estimate transmission case

for the maximum RMSE. But DDC gives communication cost savings since it needs only half of the communication rate of FC.

2.6 Summary and conclusions

This chapter addresses the state estimation problem for the cases of transmitting measurements and local estimates of multiple sensors with limited communication resources. In the case of transmitting measurements, we design a DDC scheme and an approximate MMSE estimator. In the case of transmitting local estimates of multiple sensors, we propose a DDC scheme and an optimal WLS estimator. These DDC schemes, which are based on the normalized information innovation vector, can achieve a trade-off between communication costs and estimation performance. These data fusion mechanisms can allow the estimator to extract information on the state from the absence of data, therefore improving its estimate.

CHAPTER 3

STABLE REMOTE STATE ESTIMATION AND FUSION WITH COMMUNICATION DRIVEN BY CUMULATIVE ESTIMATE INNOVATION

The designing of a metric that picks data of significance is a key ingredient for data-driven communication schemes. Common metrics include variance-based, KL divergence-based, innovation-based. Among them, innovation-based metric is acclaimed as promising [40, 41, 45, 46, 49, 60, 70, 77, 80, 87]. For example, based on the measurement innovation, [40] suggested a deterministic thresholding scheme and [41] designed a stochastic thresholding scheme. [77] proposed novel transmission strategies based on minimizing the volume of the non-transmission region and ensuring mean-square stability of the estimation error at the fusion node.

It is noteworthy that the efficient innovation-based scheme that transmits significant data for better performance still awaits thorough investigation. The current literature focuses on instant innovations. Transmission will not be triggered for a long time if the innovation at each time is not large enough individually, but these innovations can be large cumulatively. Meanwhile, the stability of the remote estimator is important. Such facts motivate our investigation reported here.

In this chapter, we design novel DDC schemes and present MMSE estimators for both single sensor and multi-sensor cases. The main contributions are as follows.

1) *Cumulative estimate innovation-based communication scheme*: The cumulative estimate innovation, a new triggering variable, is introduced to reflect the importance of data. The proposed scheme is path-dependent and has the cumulative property. It can detect oscillation and does not suffer from the problem of “no transmission for a long time”. Also, the cumulative estimate innovation is less random than the instant innovation. As such, the threshold is more easily determined for a better solution of this mixed estimation-communication problem. Moreover, what is transmitted contains more information for estimation than not transmitted, so the degradation of estimation quality due to reduced communication is minimal. What’s more, for the remote center, “no trans-

mission" is informative, and this information is used optimally to improve estimation performance.

2) *Stable state estimator and fuser*: For the single-sensor case and the multi-sensor case, respectively, we derive the MMSE estimator and the MMSE-optimal WLS fuser. The estimators are optimal under well justified assumptions and have guaranteed stability. It is shown that the expected norms of the MSE matrices are bounded, and upper bounds are derived. We also give conditional probabilities of a future transmission and conditional probabilities based on the elapsed time since the latest transmission.

The chapter is organized as follows. Section 3.2 formulates the remote state estimation problem. Section 3.3 presents our DDC scheme. Section 3.4 derives the corresponding MMSE estimator. Section 3.5 analyzes stability. Section 3.6 extends the study into the multi-sensor case. Section 3.7 presents a simulation study. Summary and conclusions are provided in Section 3.8.

3.1 Problem formulation

Consider the following discrete-time stochastic linear time-invariant system

$$x_{k+1} = Ax_k + w_k \tag{3.1}$$

$$y_k = Cx_k + v_k, k \in \mathbb{Z} \tag{3.2}$$

where $x_k \in \mathbb{R}^n$ is the system state, $A \in \mathbb{R}^{n\times n}$ ($n \in \mathbb{N}$) is the system matrix, $y_k \in \mathbb{R}^m$ ($m \in \mathbb{N}$) is the measurement, $C \in \mathbb{R}^{m\times n}$ is the output matrix, w is zero-mean Gaussian white process noise with covariance $Q > 0$, and v is zero-mean Gaussian white measurement noise with covariance $R > 0$. The initial state is $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$.

Assumption 3.1 *The initial state x_0, the process noise sequence $\langle w_k\rangle$, and the measurement noise sequence $\langle v_k\rangle$ are independent.*

Assumption 3.2 *(A, C) is observable.*

Suppose that the sensor generating measurements according to (3.2) has resources to find optimal local state estimates $\hat{x}_k^L$ based on model (3.1)-(3.2). It is well known that for a linear Gaussian system (3.1)-(3.2), the Kalman filter is optimal in the MMSE sense. Consequently, for any sensor with estimate $(\hat{x}_{k-1}^L, P_{k-1}^L)$, the local estimation is

$$(\hat{x}_{k|k-1}^L, P_{k|k-1}^L, \hat{x}_k^L, P_k^L, K_k^L) = \mathbf{KF}(A, C, Q, R, y_k) \tag{3.3}$$

where $\mathbf{KF}$ is defined in (1.21).

The local estimates $\hat{x}_k^L$ are coded into data packets and transmitted to the estimator at the remote center via a communication channel. We aim to jointly design a communication scheme and a remote estimator so as to reduce the communication rate while having good estimation quality. In this chapter, we focus on data/event based scheme to decide when and what data should be transmitted. The system architecture is shown in Fig. 3.1.

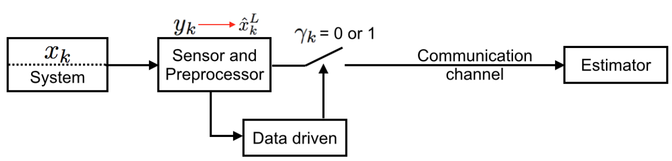


Figure 3.1: The system architecture

Since the innovation is the key to estimation, we consider a scheme based on the innovation with a binary decision variable γ_k for the sensor to transmit (or not) to the remote center. Only data whose significance is more than a pre-specified threshold will be transmitted (otherwise the innovation will be accumulated until it is large enough to be transmitted) to save communication resources. At each time k, the information about $\hat{x}_k^L$ available at the remote center is denoted by $\mathcal{I}_k$. Given $\{\mathcal{I}\}_0^k$, the MMSE estimator at the remote center is

$$\hat{x}_k = \mathbb{E}[x_k|\{\mathcal{I}\}_0^k] \tag{3.4}$$

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)'|\{\mathcal{I}\}_0^k]. \tag{3.5}$$

In this chapter, we aim to answer the following questions.

1) For a DDC scheme, how to quantify the importance of data using the innovation?

2) Given such a DDC scheme, what is the MMSE estimator?

3) Is the MMSE estimator stable?

4) How to deal with the above problems for the multi-sensor case?

Remark 3.1 *We consider a perfect communication channel. A complete analysis of the remote state estimation with DDC involving packet drop/delay and/or other issues is beyond the scope of this chapter and will be considered elsewhere.*

3.2 Scheme of communication driven by cumulative estimate innovation

In this section we propose a DDC scheme based on a cumulative estimate innovation for the single-sensor case.

Define the (instant) estimate innovation z_k at time k as

$$\begin{aligned} z_k &= \mathbb{E}[x_k|\{y\}_0^k] - \mathbb{E}[x_k|\{y\}_0^{k-1}] \\ &= \hat{x}_k^L - \hat{x}_{k|k-1}^L = K_k^L \tilde{y}_k, \end{aligned} \tag{3.6}$$

where $\tilde{y}_k = y_k - C\hat{x}_{k|k-1}^L$ is the measurement innovation. Let $\Delta P_k^L = P_{k|k-1}^L - P_k^L$.

Lemma 3.1 *The estimate innovation sequence $\{z\}_0^k$ has the following properties:*
1) $z_k \sim \mathcal{N}(\mathbf{0}, \Delta P_k^L)$.
2) $\{z\}_0^k$ is an independent sequence.

1) z_k is Gaussian because it is a linear combination of the Gaussian $y_0, y_1, \ldots, y_k$. Since $\mathbb{E}[\tilde{y}_k] = 0$, letting $\mathbb{E}[\tilde{y}_k \tilde{y}_k'] = S_k$, we have

$$\mathbb{E}[z_k] = K_k^L \mathbb{E}[\tilde{y}_k] = \mathbf{0}, \tag{3.7}$$

$$\mathbb{E}[z_k z_k'] = K_k^L S_k (K_k^L)' = P_{k|k-1}^L - P_k^L. \tag{3.8}$$

Part 2 follows directly from the independence of the measurement innovation sequence [37] since z_k is proportional to the measurement innovation.

Lemma 3.2 *$\Delta P_k^L \geq 0$, $\text{rank}(\Delta P_k^L) = q$, where $q = \text{rank}(C)$.*

Since $Q, R > 0$, we have $P_{k|k-1}^L > 0$ and $(CP_{k|k-1}^L C' + R)^{-1} > 0$, and therefore, $(CP_{k|k-1}^L C' + R)^{-1} = L_k' L_k$, where L_k is full rank. We have

$$P_{k|k-1}^L - P_k^L = K_k^L S_k (K_k^L)' \tag{3.9}$$

$$= P_{k|k-1}^L C'(CP_{k|k-1}^L C' + R)^{-1} CP_{k|k-1}^L \tag{3.10}$$

$$= (L_k CP_{k|k-1}^L)' L_k CP_{k|k-1}^L \geq 0. \tag{3.11}$$

Thus, rank$(P_{k|k-1}^L - P_k^L) =$ rank$(C) = q$.

Since $\Delta P_k^L \geq 0$, there exists a unitary matrix $U_k \in \mathbb{R}^{n \times n}$ such that

$$U_k' \Delta P_k^L U_k = \text{diag}(\Lambda_k, \mathbf{0}_{q \times (n-q)}) \tag{3.12}$$

where $\Lambda_k = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_q)$ and $\lambda_1, \lambda_2, \cdots, \lambda_q$ are the positive eigenvalues of ΔP_k^L. Let $F_k = [(\Lambda_k)^{-\frac{1}{2}}, \mathbf{0}_{q \times (n-q)}] U_k'$. Then define the normalized instant estimate innovation vector as

$$\epsilon_k = F_k z_k. \tag{3.13}$$

For each time k, define the latest transmission time

$$t_k = \max\{t \in \mathbb{Z} : t < k, \gamma_t = 1\}. \tag{3.14}$$

Thus, the elapsed time τ_k is

$$\tau_k = k - t_k, \;\; \tau_k \in \{1, 2, \ldots, k\} \tag{3.15}$$

which denotes, at a generic time k, the time elapsed since the latest transmission. When the time index is clear from the context, we will write τ_k as τ and t_k as t for simplicity. Then, it follows from (1.16), (1.18), (3.6) that $\hat{x}_k^L$ can be represented as

$$\begin{aligned}\hat{x}_k^L &= A^\tau \hat{x}_t^L + A^{\tau-1} z_{t+1} + \cdots + z_k \\ &= A^\tau \hat{x}_t^L + A^{\tau-1} F_{t+1}^\dagger \epsilon_{t+1} + \cdots + F_k^\dagger \epsilon_k \\ &= A^\tau \hat{x}_t^L + A^F \epsilon_{t+1,k} \qquad (3.16)\end{aligned}$$

where

$$F_k^\dagger = U_k[(\Lambda_k)^{\frac{1}{2}}, \mathbf{0}_{q\times(n-q)}]' \qquad (3.17)$$

$$\epsilon_{t+1,k} = [\epsilon_{t+1}', \epsilon_{t+2}', \ldots, \epsilon_k']' \qquad (3.18)$$

$$A^F = [A^{\tau-1} F_{t+1}^\dagger, \ldots, A F_{k-1}^\dagger, F_k^\dagger]. \qquad (3.19)$$

Obviously, $A^F \epsilon_{t+1,k}$ contains all innovations in the local estimate at the current time k since the latest transmission.

We propose a DDC scheme based on the cumulative estimate innovation $\epsilon_{t+1,k}$ and the decision variable

$$\gamma_k = \begin{cases} 1 & \text{if } ||\epsilon_{t+1,k}||^2 > \delta \\ 0 & \text{if } ||\epsilon_{t+1,k}||^2 \leq \delta \end{cases} \qquad (3.20)$$

where $||\epsilon_{t+1,k}||^2$ is a statistic and $\delta > 0$ is a pre-specified threshold. Denote

$$\tilde{\epsilon}_{k,\tau} = \mathbb{E}[x_k | \{y\}_0^k] - \mathbb{E}[x_k | \{y\}_0^t] \qquad (3.21)$$

$$= \hat{x}_k^L - A^\tau \hat{x}_t^L = A^F \epsilon_{t+1,k}. \qquad (3.22)$$

If $\gamma_k = 1$, then the sensor sends $\tilde{\epsilon}_{k,\tau}$, which is the cumulative estimate innovation at time k since the latest transmission. Also, information "no data transmission" at time k, corresponding to $||\epsilon_{t+1,k}||^2 \leq \delta$, can be and should be used to improve estimation performance.

Note that $||\epsilon_{t+1,k}||^2 = ||\epsilon_{t+1}||^2 + ||\epsilon_{t+2}||^2 + \cdots + ||\epsilon_k||^2$. A flowchart of the data-driven communication scheme is shown in Fig. 3.2.

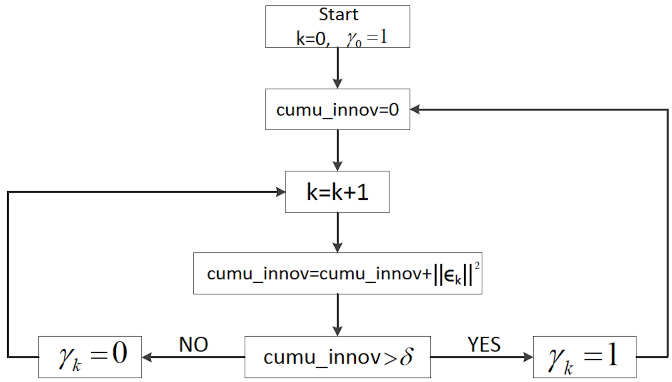


Figure 3.2: A flowchart of DDC scheme

Remark 3.2 *In our scheme, the positive and the negative parts of z_{t+1}, $\ldots$, z_k in (3.20) do not cancel each other. It is path-dependent and has the cumulative property. The cumulative statistic $||\epsilon_{t+1,k}||^2$ is non-decreasing in the elapsed time. It can, without any ad hoc constraint, naturally avoid the unfavorable situation of "no triggering for a long time even with a large cumulation". It tailors to the individual path, rather than say, with some additional conditions which are data/case independent, as in [78] and [95]. Also, the cumulative estimate innovation is less random since it involves more terms than the highly random instant innovation. With it, the threshold can be determined more easily, directly, and reliably for a better solution of this mixed estimation-communication problem.*

Remark 3.3 *$\epsilon_{t+1,k}$ as a stack of estimate innovations from time $t+1$ to k is essentially an estimate innovation since the latest transmission. Information "no data transmission" at time k, corresponding to $||\epsilon_{t+1,k}||^2$ not exceeding a threshold, can also be used optimally to improve estimation performance.*

3.3 Remote MMSE estimation

This section shows how to design a remote MMSE estimator uses single-sensor data under our communication scheme.

Under our communication scheme, the available data at the remote center at time k is $\{\mathcal{I}\}_0^k$, where $\mathcal{I}_k = \{1, \hat{x}_k^L - A^\tau \hat{x}_t^L\}$ or $\{0\}$. Here $\{1\}$ or $\{0\}$ indicates $\gamma_k = 1$ or 0.

Let the region $\Omega_k = \{\epsilon_{t+1,k} \in \mathbb{R}^{\tau q} : ||\epsilon_{t+1,k}||^2 \leq \delta\}$ carry information about the sensor data with $\gamma_k = 0$. Note that if $\gamma_k = 0$, all the remote center knows is $\epsilon_{t+1,k} \in \Omega_k$. Obviously, $\epsilon_{t+1,k} \in \Omega_k \Rightarrow \epsilon_{t+1,k-1} \in \Omega_{k-1}, \cdots, \epsilon_{t+1,t+1} \in \Omega_{t+1}$, i.e., $\epsilon_{t+1,i} \in \Omega_i$ is sufficient but not necessary for $\epsilon_{t+1,j} \in \Omega_j (i \geq j \geq t+1)$ since $||\epsilon_{t+1,i}||^2 \geq ||\epsilon_{t+1,j}||^2$. Then $\mathcal{I}_k$ can be represented as

$$\mathcal{I}_k = \begin{cases} \{1, \hat{x}_k^L - A^\tau \hat{x}_t^L\} & \text{if } \gamma_k = 1 \\ \{0, \Omega_k\} & \text{if } \gamma_k = 0. \end{cases} \tag{3.23}$$

Thus, if $\gamma_k = 0$, then $\{\mathcal{I}\}_0^k = \{\{\mathcal{I}\}_0^t, \Omega_{t+1}, \ldots, \Omega_{k-1}, \Omega_k\} = \{\{\mathcal{I}\}_0^t, \Omega_k\}$ since $\epsilon_{t+1,k} \in \Omega_k$ implies $\epsilon_{t+1,k-1} \in \Omega_{k-1}, \cdots, \epsilon_{t+1,t+1} \in \Omega_{t+1}$.

Lemma 3.3

$$\mathbb{E}[\epsilon_{t+1,k} | \{\mathcal{I}\}_0^t, \Omega_k] = \mathbf{0}, \tag{3.24}$$

$$\mathbb{E}[\epsilon_{t+1,k} (\epsilon_{t+1,k})' | \{\mathcal{I}\}_0^t, \Omega_k] = \beta(\tau, \delta) I_{\tau q} \tag{3.25}$$

where $\beta(\tau, \delta) = \frac{Pr(\chi^2_{\tau q+2} \leq \delta)}{Pr(\chi^2_{\tau q} \leq \delta)}$ and $I_{\tau q}$ is the identity matrix of size τq.

Proof: Unconditionally, $\epsilon_k \sim \mathcal{N}(\mathbf{0}, I_q)$ according to Lemma 3.1 and (3.13). Also, $\epsilon_{t+1}, \epsilon_{t+2}, \ldots, \epsilon_k$ are independent. Thus, unconditionally, $\epsilon_{t+1,k} \sim \mathcal{N}(\mathbf{0}, I_{\tau q})$.

It is clear that $\epsilon_{t+1,k}$ and $\{\mathcal{I}\}_0^t$ are independent. When $\gamma_k = 0$, i.e., given $\{\mathcal{I}\}_0^t$ and Ω_k, $\epsilon_{t+1,k}$ is "Gaussian" but truncated to the region Ω_k, and thus, by definition, it is truncated Gaussian [109]. According to Lemma 3 in [110], its mean and covariance are given by (3.24)-(3.25). $\square$

Lemma 3.3 gives the mean and covariance of $\epsilon_{t+1,k}$ conditioned on $\gamma_k = 0$. The covariance depends on the threshold δ and the elapsed time τ. Each entry of the covariance is less than 1 since $\beta(\tau, \delta) < 1$ for $\delta < \infty$ and can be much smaller than 1.

Lemma 3.4 *$\beta(\tau, \delta)$ increases with δ.*

Let $\frac{\tau q}{2} = g > 0$, $\frac{\delta}{2} = h > 0$.

$$\begin{aligned}\beta(\tau,\delta) &= \frac{\Pr(\chi^2_{\tau q+2} \le \delta)}{\Pr(\chi^2_{\tau q} \le \delta)} = \frac{2}{\tau q}\frac{r(\frac{\tau q}{2}+1, \frac{\delta}{2})}{r(\frac{\tau q}{2}, \frac{\delta}{2})} \\ &= \frac{1}{g}\frac{r(g+1,h)}{r(g,h)} \end{aligned} \tag{3.26}$$

$$\begin{aligned}&\frac{\partial\beta}{\partial h} = \frac{h^{g-1}e^{-h}}{gr^2(g,h)}(hr(g,h) - r(g+1,h)) \\ &= \frac{h^{2g}e^{-2h}}{gr^2(g,h)}(\frac{1}{g}M(1,g+1,h) - \frac{1}{g+1}M(1,g+2,h)) \\ &= \frac{h^{2g}e^{-2h}}{gr^2(g,h)}((\frac{1}{g} + \frac{h}{g(g+1)} + \frac{h^2}{g(g+1)(g+2)} + \dots) \\ &- (\frac{1}{g+1} + \frac{h}{(g+1)(g+2)} + \frac{h^2}{(g+1)(g+2)(g+3)} + \dots)) \\ &= \frac{h^{2g}e^{-2h}}{gr^2(g,h)}(\frac{1}{g(g+1)} + \frac{2h}{g(g+1)(g+2)} + \dots) > 0 \end{aligned} \tag{3.27}$$

where $r(\cdot,\cdot)$ is the lower incomplete gamma function, $M(\cdot,\cdot,\cdot)$ is Kummer's confluent hypergeometric function, and the second equality above holds because $r(g,h) = g^{-1}h^g e^{-h}M(1,g+1,h)$.

Theorem 3.1 *In the single-sensor case, consider the remote state estimation with our communication scheme. Under Assumption 3.1, the MMSE estimator of x_k given $\{\mathcal{I}\}_0^k$ is*

$$\hat{x}_k = A\hat{x}_{k-1} + \gamma_k(\hat{x}_k^L - A^\tau \hat{x}_t^L) \tag{3.28}$$

$$P_k = P_k^L + (1-\gamma_k)\beta(\tau,\delta)\sum_{i=0}^{\tau-1} A^i(P^L_{k-i|k-i-1} - P^L_{k-i})(A^i)'. \tag{3.29}$$

Proof: Let $p_1, p_2, \cdots, p_l$ ($p_l = t$ of (3.14)) be the time sequence at which the sensor transmits data. Then,

$$\hat{x}_{p_1} = \bar{x}_0 = \hat{x}^L_{p_1}$$

$$\hat{x}_{p_2} = A^{p_2-p_1}\hat{x}_{p_1} + (\hat{x}^L_{p_2} - A^{p_2-p_1}\hat{x}^L_{p_1}) = \hat{x}^L_{p_2}$$

$$\vdots \tag{3.30}$$

$$\hat{x}_{p_l} = A^{p_l - p_{l-1}} \hat{x}_{p_{l-1}} + (\hat{x}^L_{p_l} - A^{p_l - p_{l-1}} \hat{x}^L_{p_{l-1}}) = \hat{x}^L_{p_l}.$$

Since $p_l = t$, we have

$$\hat{x}_t = \hat{x}^L_t. \tag{3.31}$$

For $\gamma_k = 1$,

$$\hat{x}_k = A^\tau \hat{x}_t + (\hat{x}^L_k - A^\tau \hat{x}^L_t) = \hat{x}^L_k \tag{3.32}$$

$$P_k = P^L_k. \tag{3.33}$$

For $\gamma_k = 0$, by the law of total expectation [111],

$$\begin{aligned} \hat{x}_k &= \mathbb{E}[x_k | \{\mathcal{I}\}_0^k] = \mathbb{E}[x_k | \{\mathcal{I}\}_0^t, \Omega_k] \\ &= \mathbb{E}[\mathbb{E}[x_k | \{\mathcal{I}\}_0^t, \epsilon_{t+1,k}] | \{\mathcal{I}\}_0^t, \Omega_k] \\ &= \mathbb{E}[\mathbb{E}[x_k | \{\mathcal{I}\}_0^t, \epsilon_{t+1}, \ldots, \epsilon_k] | \{\mathcal{I}\}_0^t, \Omega_k]. \end{aligned} \tag{3.34}$$

Note that

$$\begin{aligned} &\mathbb{E}[x_k | \{\mathcal{I}\}_0^t, \epsilon_{t+1}, \ldots, \epsilon_k] \\ &= A^\tau \hat{x}_t + A^{\tau-1} F^{-1}_{t+1} \epsilon_{t+1} + \cdots + F^{-1}_k \epsilon_k \\ &= A^\tau \hat{x}_t + A^F \epsilon_{t+1,k}. \end{aligned} \tag{3.35}$$

By Lemma 3.3 and (3.31), we have

$$\begin{aligned} \hat{x}_k &= A^\tau \hat{x}_t + A^F \mathbb{E}[\epsilon_{t+1,k} | \{\mathcal{I}\}_0^t, \Omega_k] \\ &= A^\tau \hat{x}_t = A^\tau \hat{x}^L_t. \end{aligned} \tag{3.36}$$

According to (3.16), we have

$$\begin{aligned} x_k - \hat{x}_k &= x_k - A^\tau \hat{x}_t^L \\ &= x_k - \hat{x}_k^L + A^F \epsilon_{t+1,k}. \end{aligned} \tag{3.37}$$

Thus,

$$\begin{aligned} P_k &= \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)' | \{\mathcal{I}\}_0^t, \Omega_k] \\ &= \mathbb{E}[(x_k - \hat{x}_k^L + A^F \epsilon_{t+1,k})(x_k - \hat{x}_k^L + A^F \epsilon_{t+1,k})' | \{\mathcal{I}\}_0^t, \Omega_k] \\ &= P_k^L + A^F \mathbb{E}[\epsilon_{t+1,k}(\epsilon_{t+1,k})' | \{\mathcal{I}\}_0^t, \Omega_k](A^F)' + \mathbb{E}[(x_k - \hat{x}_k^L)(\hat{x}_k^L - A^\tau \hat{x}_t^L)' | \{\mathcal{I}\}_0^t, \Omega_k] \\ &+ \mathbb{E}[(\hat{x}_k^L - A^\tau \hat{x}_t^L)(x_k - \hat{x}_k^L)' | \{\mathcal{I}\}_0^t, \Omega_k] \\ &= P_k^L + A^F \beta(\tau, \delta)(A^F)' \qquad (3.38) \\ &= P_k^L + \beta(\tau, \delta) \sum_{i=0}^{\tau-1} A^i (P_{k-i|k-i-1}^L - P_{k-i}^L)(A^i)' \end{aligned}$$

where the fourth equality above holds because $x_k - \hat{x}_k^L$ is orthogonal to $\hat{x}_k^L$ and $A^\tau \hat{x}_t^L$.

According to (3.32) and (3.36), we have

$$\begin{aligned} \hat{x}_k &= A^\tau \hat{x}_t + \gamma_k (\hat{x}_k^L - A^\tau \hat{x}_t^L) \\ &= A \hat{x}_{k-1} + \gamma_k (\hat{x}_k^L - A^\tau \hat{x}_t^L). \end{aligned} \tag{3.39}$$

By (3.33) and (3.38), we obtain (3.29). □

Comparing with the remote estimator using all local estimates, the difference lies in the update when $\gamma_k = 0$. If $\gamma_k = 0$, the available information from the sensor is given by the predicted local estimate $A^\tau \hat{x}_t^L$. The corresponding P_k is updated as the local MSE P_k^L plus an additional term $\beta(\tau, \delta) \sum_{i=0}^{\tau-1} A^i (P_{k-i|k-i-1}^L - P_{k-i}^L)(A^i)'$.

This additional term arises due to the uncertainty in the non-transmitted data in the interval $[t+1, k]$. The $\tilde{\epsilon}_{k,\tau}$ is sent only when the cumulative estimate innovation is large enough. Therefore, at the remote center, the fact of no transmission is a piece of useful information indicating that the cumulative innovation is small enough. A small cumulative innovation suggests that the predicted local state estimate has a small error than otherwise. In this sense, each non-transmission is equivalent to sending the state prediction $A^\tau \hat{x}_t^L$ with a reduced MSE matrix. This leads to a reduction of the uncertainty in the non-transmitted local estimates, which helps reduce future estimation error at the remote center. Thus, P_k in (3.38) is smaller than $P_{k|t}^L$, which uses only the transmitted data. According to (3.38), P_k is equal to $P_{k|t}^L$ if and only if $\beta(\tau, \delta) = 1$, which occurs only as $\delta \to \infty$. For $\delta < \infty$, we have $\beta(\tau, \delta) < 1$ and thus $P_k < P_{k|t}^L$. A smaller δ also indicates that the $\hat{x}_k^L$ is closer to $A^\tau \hat{x}_t^L$, and thus the error of the predicted local estimates is smaller. This can be seen from Lemma 3.4. The eventual effect is a smaller P_k.

Remark 3.4 *In this chapter, we show that the cumulative estimate innovation has a truncated normal distribution, and derive the corresponding optimal estimator.*

3.4 Stability analysis

In this section, we analyze the stability of the proposed MMSE estimator given in Theorem 3.1. We also discuss conditional probability of a future transmission and conditional probability based on the elapsed time since the latest transmission.

Theorem 3.2 *In the single-sensor case, consider remote state estimation with our communication scheme. Under Assumptions 3.1 and 3.2, the expected induced 2-norm of the MSE matrix P_k in (3.29) is bounded, specifically,*

$$\begin{aligned}
&\mathbb{E}[||P_k||] \\
&\leq M + \mathbf{upper}(A, C, Q, R, \delta, q) \\
&= M + N \sum_{j=1}^{k_1-1} \sum_{i=0}^{j-1} a^i \boldsymbol{Pr}(\chi_{jq+2}^2 \leq \delta) + N e^{-\frac{\delta}{2}}
\end{aligned}$$

$$
\times \begin{cases} \phi_{k_2}(e\delta) & \textit{if } a = 1 \\ (1-a)^{-1}\psi_{k_2}(e\delta) & \textit{if } a < 1 \\ (a-1)^{-1}a^{\frac{-2}{q}}\psi_{k_3}(a^{\frac{2}{q}}e\delta) & \textit{if } a > 1 \end{cases} \tag{3.40}
$$

where

$$M = \sigma_{max}(C^{-1}RC^{-1'}) \tag{3.41}$$

$$N = \sigma_{max}(AC^{-1}RC^{-1'}A' + Q(Q + C^{-1}RC^{-1'})^{-1}Q) \tag{3.42}$$

$$a = \lambda_{max}(A'A) \tag{3.43}$$

$$k_1 = \textit{min}\{j \in \mathbb{N} : \Gamma_j(\delta) < 1\} \tag{3.44}$$

$$k_2 = \textit{min}\{j \in \mathbb{N} : \Gamma_j(e\delta) < 1\} \tag{3.45}$$

$$k_3 = \textit{min}\{j \in \mathbb{N} : \Gamma_j(a^{\frac{2}{q}}e\delta) < 1\} \tag{3.46}$$

$$\Gamma_j(i) = \frac{i}{jq+2} \tag{3.47}$$

$$\phi_l(i) = \sum_{j=k_1}^{l-1} j\Gamma_j(i)^{\frac{jq+2}{2}} + \frac{\Gamma_l(i)^{\frac{q+2}{2}}}{(1-\Gamma_l(i)^{\frac{q}{2}})^2} \tag{3.48}$$

$$\psi_l(i) = \sum_{j=k_1}^{l-1} \Gamma_j(i)^{\frac{jq+2}{2}} + \frac{\Gamma_l(i)^{\frac{lq+2}{2}}}{1-\Gamma_l(i)^{\frac{q}{2}}} \tag{3.49}$$

for $\forall i \in \mathbb{R}$, $\forall j \in \mathbb{N}$ *and* $\forall l \in \mathbb{N}$.

Proof: If $\gamma_k = 1$, by Lemma 4.1 of [112] and Assumption 3.2, $P_k = P_k^L \leq C^{-1}RC^{-1'}$. Since $P_{k|k-1}^L \geq Q$, we have

$$
\begin{aligned}
P_k^L &= ((P_{k|k-1}^L)^{-1} + C'R^{-1}C)^{-1} \\
&\geq (Q^{-1} + C'R^{-1}C)^{-1} \\
&= Q - Q(Q + C^{-1}RC^{-1'})^{-1}Q
\end{aligned} \tag{3.50}
$$

where the matrix inversion lemma is used.

For the $\gamma_k = 0$ case, we have

$$\begin{aligned}
&||\sum_{i=0}^{\tau-1} A^i (P^L_{k-i|k-i-1} - P^L_{k-i})(A^i)'|| \\
&\leq \sum_{i=0}^{\tau-1} ||A^i|| \cdot ||P^L_{k-i|k-i-1} - P^L_{k-i}|| \cdot ||(A^i)'|| \\
&\leq \sum_{i=0}^{\tau-1} ||A||^{2i} \cdot ||AC^{-1}RC^{-1'}A' + Q(Q + C^{-1}RC^{-1'})^{-1}Q|| \\
&\leq \sum_{i=0}^{\tau-1} a^i N
\end{aligned} \tag{3.51}$$

where Fact 9.8.41 and Proposition 9.4.9 of [113] are used.

If $\Gamma_j(\delta) \in (0,1)$, then $\Pr(\chi^2_{jq+2} \leq \delta) \leq (\Gamma_j(\delta) e^{1-\Gamma_j(\delta)})^{\frac{jq+2}{2}} = e^{-\frac{\delta}{2}} \Gamma_j(e\delta)^{\frac{jq+2}{2}}$. Thus, $\exists k_1 \in \mathbb{N}$ such that

$$\begin{aligned}
&\mathbb{E}[||P_k||] \\
&\leq M + \mathbb{E}[||\beta(\tau,\delta) \sum_{i=0}^{\tau-1} A^i (P^L_{k-i|k-i-1} - P^L_{k-i})(A^i)'||] \\
&\leq M + \sum_{j=1}^{\infty} \left[\sum_{i=0}^{j-1} a^i N \frac{\Pr(\chi^2_{jq+2} \leq \delta)}{\Pr(\chi^2_{jq} \leq \delta)} \right] \Pr(\chi^2_{jq} \leq \delta) \\
&\leq M + N \sum_{j=1}^{k_1-1} \sum_{i=0}^{j-1} a^i \Pr(\chi^2_{jq+2} \leq \delta) + N e^{-\frac{\delta}{2}} \Gamma
\end{aligned} \tag{3.52}$$

where $\Gamma = \sum_{j=k_1}^{\infty} \sum_{i=0}^{j-1} a^i \Gamma_j(e\delta)^{\frac{jq+2}{2}}$.

When $a = 1$, $\exists k_2 \in \mathbb{N}$ such that

$$\begin{aligned}
\Gamma &\leq \sum_{j=k_1}^{k_2-1} j\Gamma_j(e\delta)^{\frac{jq+2}{2}} + \sum_{j=1}^{\infty} j\Gamma_{k_2}(e\delta)^{\frac{jq+2}{2}} \\
&= \sum_{j=k_1}^{k_2-1} j\Gamma_j(e\delta)^{\frac{jq+2}{2}} + \Gamma_{k_2}(e\delta)^{\frac{q+2}{2}} \sum_{j=1}^{\infty} j(\Gamma_{k_2}(e\delta)^{\frac{q}{2}})^{j-1} \\
&\leq \sum_{j=k_1}^{k_2-1} j\Gamma_j(e\delta)^{\frac{jq+2}{2}} + \frac{\Gamma_{k_2}(e\delta)^{\frac{q+2}{2}}}{(1 - \Gamma_{k_2}(e\delta)^{\frac{q}{2}})^2}
\end{aligned} \tag{3.53}$$

where the method of differentiation is used for convergent infinite series (see Problem 144 in [114]).

When $a < 1$, $\exists k_2 \in \mathbb{N}$ such that

$$\begin{aligned}\Gamma &\le \frac{1}{1-a}\sum_{j=k_1}^{\infty}\Gamma_j(e\delta)^{\frac{jq+2}{2}} \\ &= \frac{1}{1-a}\left(\sum_{j=k_1}^{k_2-1}\Gamma_j(e\delta)^{\frac{jq+2}{2}} + \sum_{j=k_2}^{\infty}\Gamma_j(e\delta)^{\frac{jq+2}{2}}\right) \\ &\le \frac{1}{1-a}\left(\sum_{j=k_1}^{k_2-1}\Gamma_j(e\delta)^{\frac{jq+2}{2}} + \frac{\Gamma_{k_2}(e\delta)^{\frac{k_2q+2}{2}}}{1-\Gamma_{k_2}(e\delta)^{\frac{q}{2}}}\right). \end{aligned} \tag{3.54}$$

When $a > 1$, $\exists k_3 \in \mathbb{N}$ such that

$$\begin{aligned}\Gamma &\le \frac{1}{a-1}\sum_{j=k_1}^{\infty}a^j\Gamma_j(e\delta)^{\frac{jq+2}{2}} \\ &= \frac{a^{\frac{-2}{q}}}{a-1}\sum_{j=k_1}^{\infty}\Gamma_j(a^{\frac{2}{q}}e\delta)^{\frac{jq+2}{2}} \\ &\le \frac{a^{\frac{-2}{q}}}{a-1}\left(\sum_{j=k_1}^{k_3-1}\Gamma_j(a^{\frac{2}{q}}e\delta)^{\frac{jq+2}{2}} + \frac{\Gamma_{k_3}(a^{\frac{2}{q}}e\delta)^{\frac{k_3q+2}{2}}}{1-\Gamma_{k_3}(a^{\frac{2}{q}}e\delta)^{\frac{q}{2}}}\right). \end{aligned} \tag{3.55}$$

By combining (3.52)-(3.55), we obtain the upper bound. □

For the single-sensor case, Theorem 3.2 gives an upper bound on the expected induced 2-norm of the MSE matrix, which depends on the maximum singular values of $C^{-1}RC^{-1'}$, $AC^{-1}RC^{-1'}A'+Q$, and A, the threshold δ, and the rank of the output matrix q. It shows that the remote estimator based on one sensor with our DDC scheme is always stable.

Proposition 3.1 *In the single-sensor case, under our communication scheme, the conditional probability of a future transmission is given by*

$$\Pr(\gamma_{k+m+1} = 1|\gamma_k = 1) \tag{3.56}$$

$$= \begin{cases} \Pr(\chi_q^2 > \delta) & \textit{if } m = 0 \\ \int_0^\delta [1 - \Pr(\chi_q^2 < \delta - x)] f(x; mq) dx & \textit{if } m \in \mathbb{N} \end{cases}$$

where $f(x; mq)$ is the pdf of χ_{mq}^2.

Define the l-step cumulative estimate innovation as $e_l = \sum_{i=1}^{l} ||\epsilon_i||^2$. Obviously, $e_l \sim \chi_{lq}^2$. Let $\mu = \max\{l \in \mathbb{Z} : e_l \le \delta\}$. Then we have

$$\Pr(\gamma_{k+m+1} = 1 | \gamma_k = 1) = \Pr(\mu = m). \tag{3.57}$$

For $m = 0$, $\Pr(\mu = m) = \Pr(\chi_q^2 > \delta)$.

For $m \in \mathbb{N}$,

$$\begin{aligned} &\Pr(\mu = m) = \Pr(e_1 \le \delta, \ldots, e_m \le \delta, e_{m+1} > \delta) \\ &= \Pr(e_m \le \delta, e_{m+1} > \delta) \\ &= \Pr(e_m + ||\epsilon_{m+1}||^2 > \delta, e_m \le \delta) \\ &= \int_0^\delta \int_{\delta - x}^\infty f_{e_m, ||\epsilon_{m+1}||^2}(x, y) dy dx \\ &= \int_0^\delta \int_{\delta - x}^\infty f_{||\epsilon_{m+1}||^2}(y) dy f_{e_m}(x) dx \\ &= \int_0^\delta (1 - \Pr(\chi_q^2 < \delta - x)) f(x; mq) dx \end{aligned} \tag{3.58}$$

where the second equality follows from the property that $e_m \ge e_{m-1} \ge \cdots \ge e_1$ and the fifth holds because e_m and $||\epsilon_{m+1}||^2$ are independent.

Proposition 3.2 *In the single-sensor case, under our communication scheme, the probability of transmission depends on the value of τ_k and is given by*

$$Pr(\gamma_k = 1 | \tau_k = j) \tag{3.59}$$

$$= 1 - \begin{cases} Pr(\chi_n^2 \leq \delta) & \text{if } j = 1 \\ \frac{Pr(\chi_{jq}^2 \leq \delta)}{Pr(\chi_{(j-1)q}^2 \leq \delta)} & \text{if } j = 2, 3, \ldots \end{cases}$$

Define e_l as before. Obviously, $e_l \sim \chi_{lq}^2$.

For $j = 2, 3, \ldots,$

$$\begin{aligned} &\Pr(\gamma_k = 0 | \tau_k = j) \\ &= \Pr(e_j \leq \delta | e_{j-1} \leq \delta) = \frac{\Pr(e_j \leq \delta, e_{j-1} \leq \delta)}{\Pr(e_{j-1} \leq \delta)} \\ &= \frac{\Pr(e_j \leq \delta)}{\Pr(e_{j-1} \leq \delta)} = \frac{\Pr(\chi_{jq}^2 \leq \delta)}{\Pr(\chi_{(j-1)q}^2 \leq \delta)}. \end{aligned} \tag{3.60}$$

The case of $j = 1$ follows directly from the definition of γ_k in (3.20).

Remark 3.5 *According to Proposition 3.2, γ_k of our communication algorithm is probabilistically dependent on the past transmission pattern—the value of τ_k. The intuition here is that the greater the τ_k is, the more outdated the latest estimate is. Thus, it seems reasonable for the decision (whether to transmit at time k) to depend on the value of τ_k. In our scheme, $Pr(\gamma_k = 1 | \tau_k = j)$ takes a nonzero value for each finite τ_k.*

3.5 Extension to multi-sensor case

In the multi-sensor case, to estimate the system state x_k in (3.1) remotely, S sensors are used to measure x_k in the form of (3.2). Each sensor i ($i \in \{1, 2, \cdots, S\}$) provides a noisy measurement

$$y_k^i = C^i x_k + v_k^i \tag{3.61}$$

where $y_k^i \in \mathbb{R}^{m_i}$ ($m_i \in \mathbb{N}$) is the measurement, $C^i \in \mathbb{R}^{m_i \times n}$ is the output matrix, and v_k^i is zero-mean Gaussian white measurement noise with covariance R^i. Define $C = [(C^1)', (C^2)', \cdots, (C^S)']'$.

Assumption 3.3 *The initial state x_0, the process noise sequence $\langle w_k \rangle$, and the measurement noise* *sequences* $\{\langle v_k^i \rangle, i \in \{1, 2, \cdots, S\}\}$ *are independent for each sensor and pairwise uncorrelated across the sensors.*

Assumption 3.4 (A, C) *is observable, but each pair* (A, C^i) *is not necessarily observable.*

Each sensor i processes its measurements to find its local state estimate $\hat{x}_k^i$. This estimate is or is not transmitted to the fusion center depending on a decision variable γ_k^i via an independent communication channel.

According to Assumption 3.4, system (3.1) and (3.61) is not necessarily observable. In order to avoid a possible divergence in estimation error, it is convenient to focus on the observable subsystem by separating it from the unobservable part (see also [77]). By introducing a suitable nonsingular transformation $(T^i)^{-1}$, the transformed system matrices take the form

$$\begin{bmatrix} A_o^i & 0 \\ A_{21}^i & A_{\bar{o}}^i \end{bmatrix} = (T^i)^{-1} A T^i, \quad \begin{bmatrix} B_o^i \\ B_{\bar{o}}^i \end{bmatrix} = (T^i)^{-1}$$
$$\begin{bmatrix} C_o^i & 0 \end{bmatrix} = C^i T^i.$$

The observable part of system (3.1) and (3.61) is an n_i-dimensional subsystem with the state $x_k^i = B_o^i x_k$:

$$x_{k+1}^i = A_o^i x_k^i + B_o^i w_k \tag{3.62}$$
$$y_k^i = C_o^i x_k^i + v_k^i. \tag{3.63}$$

For each observable subsystem (3.62)-(3.63) with estimate $(\hat{x}_{k-1}^i, P_{k-1}^i)$, the local estimation by the Kalman filter is

$$(\hat{x}_{k|k-1}^i, P_{k|k-1}^i, \hat{x}_k^i, P_k^i, K_k^i) = \mathbf{KF}(A_o^i, C_o^i, B_o^i Q (B_o^i)', R^i, y_k^i). \tag{3.64}$$

We generalize the definition of z_k, ϵ_k, t_k, τ_k, $\epsilon_{t+1,k}$, F_k, A^F, δ, q to the multi-sensor case with

z_k^i, ϵ_k^i, t_k^i, τ_k^i, $\epsilon_{t^i+1,k}^i$, F_k^i, $A^{F,i}$, δ^i, q^i for the i-th sensor. When the time index is clear from the context, we will write t_k^i as t^i and τ_k^i as τ^i for simplicity.

Based on (3.20), a DDC scheme is defined as γ_k^i for sensor i. If $\gamma_k^i = 1$, the data $\tilde{\epsilon}_{k,\tau^i}^i \triangleq A^{F,i}\epsilon_{t^i+1,k}^i$ is transmitted to the fusion center.

We also generalize the definitions of Ω_k and $\mathcal{I}_k$ to the multi-sensor case with Ω_k^i and $\mathcal{I}_k^i$. Then the available data at the fusion center at each time k is $\mathcal{I}_k = \{\mathcal{I}_k^1, \dots, \mathcal{I}_k^S\}$. The fusion center knows the parameters of each observable subsystem.

Assumption 3.5 *The "non-transmitted data" of arbitrary sensors i and j, $\epsilon_{t^i+1,k}^i$ and $\epsilon_{t^j+1,k}^j$ ($i \neq j$), are mutually uncorrelated for any time k.*

Theorem 3.3 *In the multi-sensor case, consider the remote estimation fusion with our communication scheme. Under Assumptions 3.3 and 3.5, the optimal WLS fuser given $\{\mathcal{I}\}_0^k$ that minimizes MSE is*

$$\hat{x}_k = (H'\Sigma_k^{-1}H)^{-1}H'\Sigma_k^{-1}Y_k \tag{3.65}$$

$$P_k = (H'\Sigma_k^{-1}H)^{-1} \tag{3.66}$$

where

$$H = [(B_o^1)', \dots, (B_o^S)']' \tag{3.67}$$

$$Y = [(Y_k^1)', \dots, (Y_k^S)']' \tag{3.68}$$

$$Y_k^i = A_o^i Y_{k-1}^i + \gamma_k^i(\hat{x}_k^i - (A_o^i)^{\tau^i}\hat{x}_{t^i}^i) \tag{3.69}$$

$$\Sigma_k = \bar{\Sigma}_k + \mathit{diag}((1-\gamma_k^1)\Pi(\tau^1,\delta^1), (1-\gamma_k^2)\Pi(\tau^2,\delta^2), \dots, (1-\gamma_k^S)\Pi(\tau^S,\delta^S)) \tag{3.70}$$

$$\bar{\Sigma}_k = \begin{bmatrix} P_k^1 & \dots & P_k^{1,S} \\ \vdots & \ddots & \vdots \\ P_k^{S,1} & \dots & P_k^S \end{bmatrix} \tag{3.71}$$

$$P_k^{i,j} = (I - K_k^i C_o^i)(A_o^i P_{k-1}^{i,j}(A_o^j)' + B_o^i Q (B_o^j)')(I - K_k^j C_o^j) \tag{3.72}$$

$$\Pi(\tau^i, \delta^i) = \beta(\tau^i, \delta^i) \sum_{j=0}^{\tau^i - 1} (A_o^i)^j (P_{k-j|k-j-1}^i - P_{k-j}^i)((A_o^i)^j)'. \tag{3.73}$$

Here Y_k^i is the remote version of $\hat{x}_k^i$, just like $\hat{x}_k$ is that of $\hat{x}_k^L$ in the single-sensor case.

Proof: By Theorem 3.1, information at the fusion center about every local estimate is (3.69). We interpret Y_k^i as a measurement of the state x_k collected through the (virtual) communication channel,

$$Y_k^i = \begin{cases} B_o^i x_k + (\hat{x}_k^i - x_k^i) & \text{if } \gamma_k^i = 1 \\ B_o^i x_k + ((A_o^i)^{\tau^i} \hat{x}_{t^i}^i - x_k^i) & \text{if } \gamma_k^i = 0. \end{cases} \tag{3.74}$$

Here $\hat{x}_k^i - x_k^i$ and $(A_o^i)^{\tau^i} \hat{x}_{t^i}^i - x_k^i$ are considered as the measurement noise u_k^i for the cases of transmission and no transmission, respectively. Summing up all sensors, the available information at the fusion center can be treated as originating from the "measurement"

$$Y_k = Hx_k + u_k \tag{3.75}$$

where $u_k = [(u_k^1)', \ldots, (u_k^S)']'$, $u_k^i = (\hat{x}_k^i - x_k^i) + (1 - \gamma_k^i)((A_o^i)^{\tau^i} \hat{x}_{t^i}^i - \hat{x}_k^i)$. We have

$$\mathbb{E}[u_k^i | \{\mathcal{I}\}_0^k] = \mathbf{0} \tag{3.76}$$

$$\begin{aligned}
&\text{cov}[u_k^i | \{\mathcal{I}\}_0^k] \\
&= \mathbb{E}[(\hat{x}_k^i - x_k^i)(\hat{x}_k^i - x_k^i)' | \{\mathcal{I}\}_0^k] \\
&+ (1 - \gamma_k^i)\mathbb{E}[(\hat{x}_k^i - (A_o^i)^{\tau^i} \hat{x}_{t^i}^i)(\hat{x}_k^i - (A_o^i)^{\tau^i} \hat{x}_{t^i}^i)' | \{\mathcal{I}\}_0^k] \\
&= P_k^i + (1 - \gamma_k^i)\mathbb{E}[A^{F,i} \epsilon_{t^i+1,k}^i (A^{F,i} \epsilon_{t^i+1,k}^i)' | \{\mathcal{I}\}_0^k] \\
&= P_k^i + (1 - \gamma_k^i)\Pi(\tau^i, \delta^i)
\end{aligned} \tag{3.77}$$

where the second equality holds because $\hat{x}_k^i - (A_o^i)^{\tau^i} \hat{x}_{t^i}^i$ is uncorrelated with $\hat{x}_k^i - x_k^i$ conditioned on $\{I\}_0^k$. Based on Assumption 3.5 and[77],

$$
\begin{aligned}
&\mathrm{cov}[u_k^i, u_k^j|\{\mathcal{I}\}_0^k] = P_k^{i,j} \\
&= \mathbb{E}[(\hat{x}_k^i - x_k^i)(\hat{x}_k^j - x_k^j)'|\{\mathcal{I}\}_0^k] + (1-\gamma_k^i)(1-\gamma_k^j) \\
&\times \mathbb{E}[(\hat{x}_k^i - (A_o^i)^{\tau^i}\hat{x}_{t^i}^i)(\hat{x}_k^j - (A_o^j)^{\tau^j}\hat{x}_{t^j}^j)'|\{\mathcal{I}\}_0^k] \\
&= \mathbb{E}[(\hat{x}_k^i - x_k^i)(\hat{x}_k^j - x_k^j)'|\{\mathcal{I}\}_0^k] + (1-\gamma_k^i)(1-\gamma_k^j) \\
&\times \mathbb{E}[A^{F,i}\epsilon_{t^i+1,k}^i(A^{F,j}\epsilon_{t^j+1,k}^j)'] \\
&= \mathbb{E}[(\hat{x}_k^i - x_k^i)(\hat{x}_k^j - x_k^j)'|\{\mathcal{I}\}_0^k]. \qquad (3.78)
\end{aligned}
$$

Note that

$$\hat{x}_k^i - x_k^i = (I - K_k^i C_o^i)[A_o^i(\hat{x}_{k-1}^i - x_{k-1}^i) - B_o^i w_{k-1}] + K_k^i v_k^i \qquad (3.79)$$

and thus

$$P_k^{i,j} = (I - K_k^i C_o^i)[A_o^i P_{k-1}^{i,j}(A_o^j)' + B_o^i Q (B_o^j)'](I - K_k^j C_o^j). \qquad (3.80)$$

Since the measurement model (3.75) is linear, the WLS fuser given by (3.65) and (3.66) is the optimal WLS fuser in the MMSE sense, where Σ_k is the covariance of the virtual noise u_k. □

Our WLS fuser differs from the general WLS fuser using all sensors' local estimates in the update when $\gamma_k^i = 0$. If $\gamma_k^i = 0$, the available information from sensor i is given by the predicted local estimate $(A_o^i)^{\tau^i}\hat{x}_{t^i}^i$. The corresponding Σ_k is updated as in the general WLS fuser plus an additional term $\Pi(\tau^i, \delta^i)$ in the corresponding part.

Like the single-sensor case, P_k in (3.66) is smaller than that of the estimator which uses only the transmitted local estimates. The eventual effect of a smaller δ^i is a smaller P_k.

Remark 3.6 *Assumption 3.5 leads to discarding the second term of cov*$[u_k^i, u_k^j|\{\mathcal{I}\}_0^k]$ *to reach*

(3.78). It can be justified as follows. a) ϵ^i and ϵ^j are from different sensors. b) The measurement innovation sequence of the Kalman filter is white. c) Innovation of a different sensor is similar to innovation of the same sensor at a different time. Based on the above three points, ϵ^i and ϵ^j can be assumed to be uncorrelated. Note that the stability of our estimator is not affected by this assumption, although the concrete upper bound given below is affected by this assumption.

Theorem 3.4 *Under Assumptions 3.3 and 3.4, the expected induced 2-norm of the MSE matrix P_k in (3.66) is bounded, specifically,*

$$\mathbb{E}[||P_k||] \leq h[M + \max_{i\in\{1,2,\cdots,S\}} \boldsymbol{upper}(A_o^i, C_o^i, B_o^i Q (B_o^i)', R^i, \delta^i, q^i)] \tag{3.81}$$

*where $M = \sum_{i=1}^{S} tr((C_o^i)^{-1} R^i (C_o^i)^{-1'})$, $h = \lambda_{max}((H'H)^{-1})$, H was defined by (3.67), and **upper** was given in (3.40) under Assumptions 3.5.*

Proof: According to the collective observability in Assumption 3.4, it is easy to see that H is of full column rank, and thus H is left invertible. Then we have

$$\begin{aligned}
&\mathbb{E}[||P_k||] = \mathbb{E}[||(H'H)^{-1}H'\Sigma_k H(H'H)^{-1}||] \\
&\leq \mathbb{E}[||(H'H)^{-1}H'|| \cdot ||\Sigma_k|| \cdot ||H(H'H)^{-1}||] \\
&= \sigma_{\max}((H'H)^{-1}H') \cdot \mathbb{E}[||\Sigma_k||] \cdot \sigma_{\max}(H(H'H)^{-1}) \\
&\leq h(\mathbb{E}[||\bar{\Sigma}_k||] + \mathbb{E}[||\text{diag}((1-\gamma_k^i)\Pi(\tau^i, \delta^i))||])
\end{aligned} \tag{3.82}$$

where the first equality follows form Proposition 6.1.5 of [113], the first inequality holds because the induced 2-norm satisfies the sub-multiplicativity, the second equality follows from the definition of the induced 2-norm, and the second inequality holds because a matrix norm satisfies the triangle inequality. According to Proposition 9.2.4 of [113] and the definition of the induced 2-norm and the Schatten norm, we have

$$||\bar{\Sigma}_k|| = ||\bar{\Sigma}_k||_{\sigma_\infty} \leq ||\bar{\Sigma}_k||_{\sigma_1} = \text{tr}(\bar{\Sigma}_k) \tag{3.83}$$

and thus,

$$
\begin{aligned}
&\mathbb{E}[||P_k||] \leq h(\mathbb{E}[||\bar{\Sigma}_k||_{\sigma_1}] + \mathbb{E}[||\text{diag}((1-\gamma_k^1)\Pi(\tau^1,\delta^1),\\
&(1-\gamma_k^2)\Pi(\tau^2,\delta^2),\ldots,(1-\gamma_k^S)\Pi(\tau^S,\delta^S))||])\\
&= h(\sum_{i=1}^{S}\text{tr}(P_k^i) + \max_{i\in\{1,2,\cdots,S\}}\mathbb{E}[||(1-\gamma_k^i)\Pi(\tau^i,\delta^i)||])\\
&\leq h(M + \max_{i\in\{1,2,\cdots,S\}}\mathbb{E}[||(1-\gamma_k^i)\Pi(\tau^i,\delta^i)||])
\end{aligned} \tag{3.84}
$$

where the first equality holds because the eigenvalues of a block diagonal matrix are simply those of the blocks, and the second inequality holds due to Lemma 4.1 of [112] and (A_o^i, C_o^i) being observable. Using the upper bound of $\mathbb{E}[||(1-\gamma_k^i)\Pi(\tau^i,\delta^i)||]$ given by Theorem 3.2, we obtain the upper bound of $\mathbb{E}[||P_k||]$. □

It shows that the remote fuser based on multiple sensors with our DDC scheme guarantees stability.

In the multi-sensor case, the conditional probability of a future transmission Pr$(\gamma_{k+m+1}^i = 1|\gamma_k^i = 1)$ and the conditional probability based on the elapsed time since the latest transmission Pr$(\gamma_k^i = 1|\tau_k^i = j)$ of sensor i are similar to those of the single sensor case: we only need to change q and δ to q^i and δ^i.

3.6 Illustrative examples

This section presents examples to illustrate the results in this chapter and discusses properties of the communication schemes and the remote estimators for the single-sensor and multi-sensor cases.

The following six communication schemes were compared for both cases.

1) Our proposed data-driven communication based on the cumulative innovation (DDC-CI): each sensor i transmits data based on (3.20). .

2) Data-Driven Estimate Transmission (DDET) of [77]: each sensor i transmits data if $||\hat{x}_k^i - \hat{x}_{k|t}^i||_{W_k^i} > \delta^i$ holds.

3) Data-Driven Measurement Transmission (DDMT) of [77]: each sensor i transmits data if $||y_k^i - \hat{y}_{k|t}^i||_{W_k^i} > \delta^i$ holds.

4) Periodic communication (PC): each sensor i periodically transmits data at rate α^i.

5) Stochastic communication (SC): each sensor i transmits data at rate α^i at random times.

6) Full communication (FC): each sensor i transmits data at each time (i.e., at full rate 1).

The communication rates α^i for the thresholds δ^i of DDC-CI, DDET and DDMT are obtained from 1000 Monte Carlo runs.

3.6.1 One sensor case

We consider an example of tracking a target taking a (nearly) constant-turn (CT) in two-dimensional space [115]. The state of its discrete-time CT model is given by $x = [\mathrm{x}, \dot{\mathrm{x}}, \mathrm{y}, \dot{\mathrm{y}}]$, and the position is measured by a sensor. The system and measurement models are characterized by the following matrices

$$A = \begin{bmatrix} 1 & \frac{\sin\omega T}{\omega} & 0 & -\frac{1-\cos\omega T}{\omega} \\ 0 & \cos\omega T & 0 & -\sin\omega T \\ 0 & \frac{1-\cos\omega T}{\omega} & 1 & \frac{\sin\omega T}{\omega} \\ 0 & \sin\omega T & 0 & \cos\omega T \end{bmatrix} \tag{3.85}$$

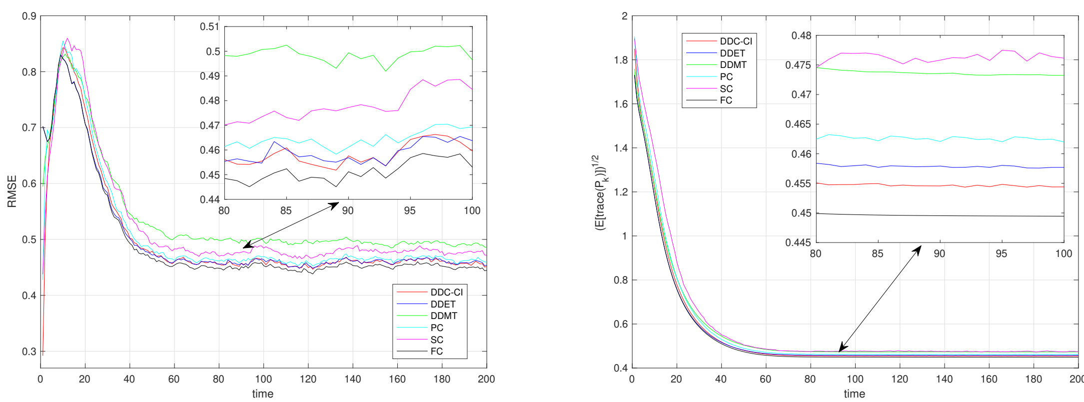
$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.86}$$

where $T = 0.1s$, $\omega = 6°/s$ is the constant turn rate, and

$$Q = q\times$$

$$
\begin{bmatrix}
\frac{2(\omega T - \sin\omega T)}{\omega^3} & \frac{1-\cos\omega T}{\omega^2} & 0 & \frac{\omega T - \sin\omega T}{\omega^2} \\
\frac{1-\cos\omega T}{\omega^2} & T & -\frac{\omega T - \sin\omega T}{\omega^2} & 0 \\
0 & -\frac{\omega T - \sin\omega T}{\omega^2} & \frac{2(\omega T - \sin\omega T)}{\omega^3} & \frac{1-\cos\omega T}{\omega^2} \\
\frac{\omega T - \sin\omega T}{\omega^2} & 0 & \frac{1-\cos\omega T}{\omega^2} & T
\end{bmatrix}
\tag{3.87}
$$

the spectral density $q = 0.01$, $R = I_2$, $P_0 = I_4$, $x_0 = [1, 10, 1, 10]'$.



(a) RMSE

(b) $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$

Figure 3.3: RMSE and the corresponding $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$ for remote estimates of DDC-CI, DDET[77], DDMT[77], PC, SC at $\alpha = 0.4$ and FC at $\alpha = 1$

Fig. 3.3 depicts root mean square error (RMSE) curves over 1000 Monte Carlo runs and the corresponding $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$ of the remote estimates of DDC-CI, DDET, DDMT, PC, SC at $\alpha = 0.4$ and FC at $\alpha = 1$. Fig. 3.3 shows that DDC-CI has a slightly lower RMSE and $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$ than those of DDET. Simulations also show that DDC-CI has a smaller error than DDMT, PC and SC. Therefore, our scheme slightly outperforms DDET and outperforms DDMT, PC and SC by using the same communication rate. Compared with FC, our DDC-CI estimator in the steady-state saves 60% of the communication costs at the price of an increase in estimation error by 1% (since 0.455/0.45 $\approx$ 1.01). Note the underperformance of DDMT compared with other schemes can be partly attributed to its transmission of measurements rather than local estimates. Each local estimate sums up all information about previous measurements.

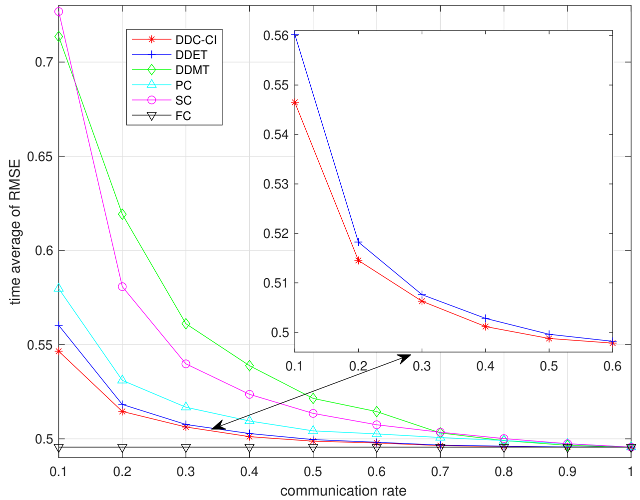


Figure 3.4: Time averaged RMSE for the six communication schemes vs. communication rate

For DDC-CI, DDET, DDMT, PC, SC and FC, Fig. 3.4 plots the time average of RMSE over 1000 Monte Carlo runs versus the communication rate. For all communication rates, DDC-CI have error smaller than DDET, DDMT, PC and SC. Our scheme performs better than these four schemes at the same communication rates. The estimation errors of the five communication schemes decrease as the communication rate increases (i.e., the threshold decreases). Simulations also show that the error of DDC-CI are very close to that of FC even at a moderate communication rate of 0.6.

3.6.2 Multi-sensor case

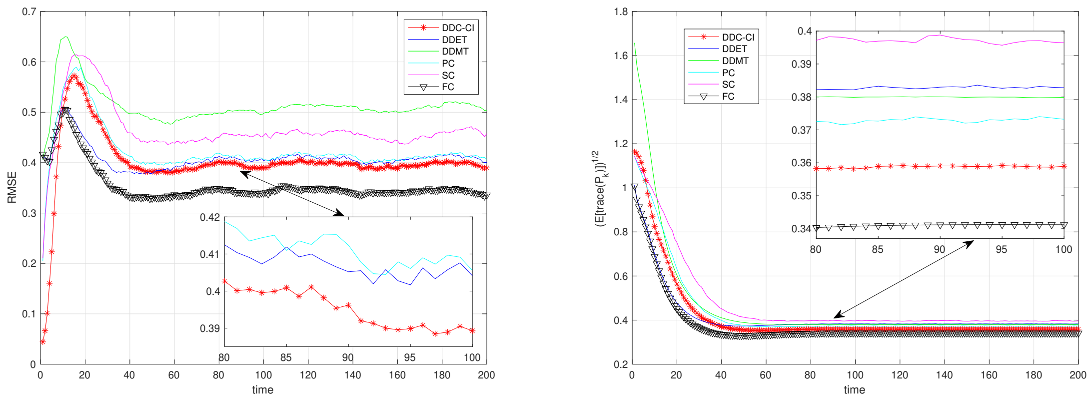
In this case, three types of sensors were used to measure the state x_k, with the following parameters:

$$C^i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, R^i = I_2, P_0^{i,j} = 0.1 I_2 \; if \; i, \; j \in \{1, 2, 3\}$$

$$C^i = [1, 0, 0, 0], \; R^i = 1, \; P_0^{i,j} = 0.1 \; if \; i, \; j \in \{4, 6, 8\}$$

$$C^i = [0, 0, 1, 0], \; R^i = 1, \; P_0^{i,j} = 0.1 \; if \; i, \; j \in \{5, 7, 9\}$$

where $i \neq j$.



(a) RMSE

(b) $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$

Figure 3.5: RMSE and the corresponding $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$ for remote estimates of DDC-CI, DDET[77], DDMT[77], PC, SC at $\alpha = 0.1$ and FC at $\alpha = 1$

Fig. 3.5 depicts RMSE curves over 1000 Monte Carlo runs and the corresponding $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$ of the estimates of DDC-CI, DDET, DDMT, PC, SC using three sensors at $\alpha = 0.1$ and FC at $\alpha = 1$. Fig. 3.5 shows that DDET has a slightly higher RMSE and $(\mathbb{E}[tr(P_k)])^{\frac{1}{2}}$ than those of DDC-CI. Also, DDC-CI has a smaller error than DDMT, PC and SC. Therefore, our scheme has the best performance among all these five schemes at the same communication rate. Compared with FC, our DDC-CI fuser in the steady-state saves 90% of the communication costs at the price of an increase in estimation error by approximately 15% (since 0.39/0.34 $\approx$ 1.15).

For DDC-CI, DDET, DDMT, PC and SC at $\alpha = 0.1$, Fig. 3.6 depicts the time average of RMSE over 1000 Monte Carlo runs versus the number of sensors. With respect to RMSE, DDC-CI exhibits a slightly smaller error than DDET for all cases. Simulations also show that, for the same number of sensors, DDC-CI always has a smaller error than DDMT, PC and SC. The estimation errors of all the five communication schemes decrease as the number of sensors increases, as expected. Fig. 3.7 shows that the time average of $\mathbb{E}[||P_k||]$ of DDC-CI is indeed upper bounded by the bound given in Theorem 3.4. Although the bound does not appear tight, it is a bound for all possible cases.

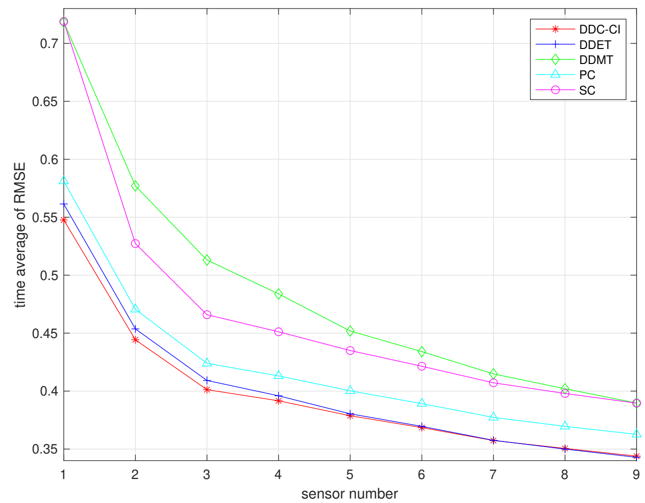


Figure 3.6: Time averaged RMSE for the five communication schemes vs. number of sensors

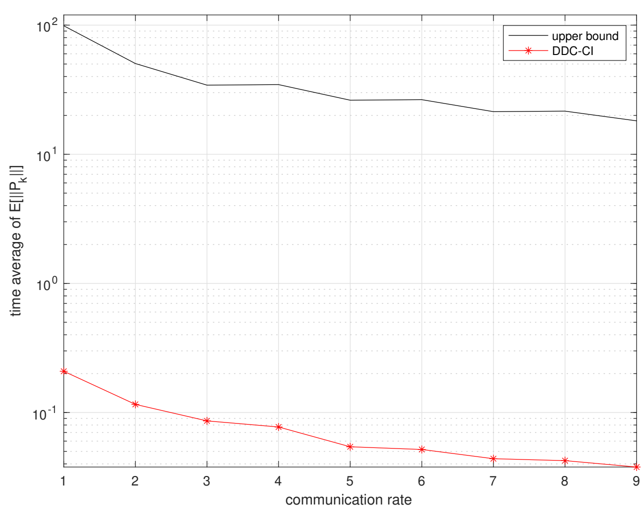


Figure 3.7: Time averaged $\mathbb{E}[||P_k||]$ for DDC-CI and its upper bound vs. number of sensors

3.7 Summary and conclusions

This chapter has studied the remote state estimation problem with limited communication resources. For the single-sensor case, we have proposed a new DDC scheme that decides whether or not to send data based on the cumulative estimate innovation since the latest transmission, and we have presented the corresponding MMSE estimator. For the multi-sensor case, we have presented a DDC scheme as an extension of the single-sensor case, and an optimal WLS fuser. By theoretical analysis, these remote estimators have been found to have a guaranteed stability in that their expected norms of the MSE matrices are upper bounded. For both cases, the conditional probabilities of a future transmission and conditional probabilities based on the elapsed time since the latest transmission are given.

The proposed communication schemes send data only when the cumulative estimate innovation since the latest transmission exceeds a pre-specified threshold, and so choosing an appropriate threshold can balance communication cost and estimation performance. The proposed estimator and fuser take advantage of the fact that a small cumulative innovation is implied if data is not transmitted. Each event of no data transmission is theoretically equivalent to sending the prediction with a reduced error covariance.

Our simulation results illustrate the advantage of our data-driven communication schemes and the associated optimal estimators over DDMT in [77], periodic communication and purely random communication. It also shows that our schemes slightly outperforms DDET in [77].

CHAPTER 4
OPTIMAL STATE ESTIMATION WITH HYBRID DATA-TIME-DRIVEN COMMUNICATION

4.1 Introduction

The mainstream literature on data-driven estimators focuses on optimal estimation instead of optimal communication schemes. To quantitatively account for communication resources, a cohort of researchers deal with the JDE problem to optimally calibrates the tradeoff between estimation quality and communication expenses. [92, 93] studied a scalar system with a limited number of transmissions. This work was extended in [94] to vector systems, in which a suboptimal solution is given. [96] considered an optimization problem which minimizes the average mean square state estimation error plus the transmission cost with respect to the transmission probability. [97] considered the joint optimal design of state estimator and event-trigger over a finite horizon, and proposed an iterative algorithm that alternates optimizing state estimator and event-trigger. [91] introduced a distributed greedy heuristic to minimize the weighted function of the network energy consumption and the number of transmissions, subject to constraints on the estimator performance. [98] considered event-triggered estimation with packet drops and showed that a threshold policy is optimal in the sense that it minimizes a convex combination of the expected error covariance and expected energy usage. [95] formulated the problem as an optimization problem with the average sending rate constraint and found a sub-optimal solution via generalized geometric programming optimization techniques. [99] studied optimal event-triggered estimation problem under a stochastic thresholding scheme.

In this chapter, we consider design of data-driven communication scheme and derive corresponding estimator to quantitatively balance system performance and limited communication resource. Our main contributions are as follows.

1) We propose a hybrid data-time-driven communication scheme based on the cumulative estimate innovation, time-varying thresholds and the maximum elapsed time and derive the MMSE estimator accordingly. We have proved, with due mathematical rigor, that the estimation error matrix—conditioned on the information carried by the entire "no data transmission" duration, does not exceed the one conditioned only on the information carried by the last "no data transmission" timestamp. We further give a workable tight upper bound for the estimation error matrix.

2) To quantify the tradeoff between estimation quality and communication expenses, we define an optimization criterion based on the expected total discounted cost—including estimation error and weighted communication cost—over the infinite horizon. We are able to ascertain a highly nontrivial finding—the stochastic linear system under a cumulative data communication scheme is, through constructing a state vector by stacking up certain system parameters, representable as a Markov Decision Process (MDP) problem. An iterative algorithm is proposed to find the optimal policy and the optimal cost. We also extend our results to the optimal average cost problem.

3) We prove a degree of freedom (DOF) reduction/degeneracy property that is peculiar to the optimal triggering policy. That is, the optimal thresholding decision is exclusively dependent on the tally of the elapsed time since the last transmission. No other state variables are needed in determining the threshold value at the immediate next decision point.

The chapter is organized as follows. Section 4.2 formulates the remote state estimation problem and the optimization problem. Section 4.3 presents our hybrid driven communication scheme. Section 4.4 derives the MMSE estimator and gives an upper bound for the trace of the MSE matrices. Section 4.5 reformats the optimization problem as an MDP problem and gives the optimal solution. Section 4.6 presents a simulation study.

4.2 System model

Consider the following discrete-time stochastic linear time-invariant system

$$x_{k+1} = Ax_k + \omega_k \tag{4.1}$$

$$y_k = Cx_k + \nu_k, k \in \mathbb{Z} \tag{4.2}$$

where $x_k \in \mathbb{R}^n$ is the system state, $A \in \mathbb{R}^{n\times n}$ ($n \in \mathbb{N}$) is the system matrix, $y_k \in \mathbb{R}^m$ ($m \in \mathbb{N}$) is the measurement, $C \in \mathbb{R}^{m\times n}$ is the output matrix, ω is zero-mean Gaussian white process noise with covariance Q, and ν is zero-mean Gaussian white measurement noise with covariance R. The initial state is $x_0 \sim \mathcal{N}(\bar{x}_0, P_0)$.

Assumption 4.1 *The initial state x_0, the process noise sequence $\langle w\rangle$, and the measurement noise sequence $\langle v\rangle$ are independent.*

Assumption 4.2 *(A, C) is observable.*

Suppose that the sensor generating measurements according to (4.2) has resources to find optimal local state estimates $\hat{x}_k^L$ based on model (4.1)–(4.2). It is well known that for a linear Gaussian system (4.1)–(4.2), the Kalman filter is optimal in the MMSE sense. Consequently, for any sensor with estimate $(\hat{x}_{k-1}^L, P_{k-1}^L)$, the local estimation $(\hat{x}_{k|k-1}^L, P_{k|k-1}^L, \hat{x}_k^L, P_k^L, K_k^L)$ is

$$(\hat{x}_{k|k-1}^L, P_{k|k-1}^L, \hat{x}_k^L, P_k^L, K_k^L) = \mathbf{KF}(A, C, Q, R, y_k), \tag{4.3}$$

where $\mathbf{KF}$ is defined in (1.21).

The local estimates $\hat{x}_k^L$ are coded into data packets and transmitted to the estimator at the remote center via a communication channel. We aim to design a communication policy $\pi^\gamma = \{\gamma_1, \dots, \gamma_k, \dots\}$ to balance communication cost and estimation quality. Here a binary decision variable $\gamma_k = 1$ is used to indicate that the sensor transmits data at time k and $\gamma_k = 0$ otherwise. Several communication schemes have been devised for this purpose. In this chapter, attention will be restricted to a type of hybrid data-time-driven strategies in which the transmission is driven predominantly by data (the cumulative estimate innovation), while, in some rare circumstances, also by time (the maximum elapsed time).

At each time k, the information about x_k available at the remote center is denoted by $\mathcal{I}_k$. Given

$\{\mathcal{I}\}_0^k$, the MMSE estimation at the remote center is

$$\hat{x}_k = \mathbb{E}[x_k|\{\mathcal{I}\}_0^k] \tag{4.4}$$

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)'|\{\mathcal{I}\}_0^k]. \tag{4.5}$$

To quantitatively make a the trade-off between communication consumption and estimation quality, the cost for each time k is formulated as $\text{tr}(P_k) + \kappa\gamma_k$, where the constant κ is a weighted cost paid per unit of time when data is transmitted. In order to take into account the long-term effect, we consider the expected total discounted cost over the infinite horizon which is defined as follows.

Problem 4.1

$$\min_{\pi^\gamma} \lim_{T\to\infty} \mathbb{E}^{\pi^\gamma}\left[\sum_{k=1}^{T} \alpha^{(k-1)}(\text{tr}(P_k) + \kappa\gamma_k)\right] \tag{4.6}$$

where $0 \leq \alpha < 1$ is the discount factor. The motivation of studying discounted problems comes mainly from a plethora of applications, where costs in the future bear less importance than the current one.

In the reminder of the chapter, we aim to answer the following questions.

1) How to design a hybrid data-time-driven communication scheme?

2) What is the MMSE estimator under our hybrid data-time-driven communication scheme? How to obtain an upper bound of the MSE matrices?

3) How to solve the above optimization problem under our hybrid data-time-driven communication scheme?

4.3 Hybrid data-time-driven communication

In this section we propose a hybrid driven communication scheme based on the cumulative estimate innovation, time-varying thresholds and the maximum elapsed time. We give conditional probabilities based on the elapsed time since the latest transmission in analytical expression.

Define the (instant) estimate innovation z_k as

$$
\begin{aligned}
z_k &= \mathbb{E}[x_k|\{y\}_0^k] - \mathbb{E}[x_k|\{y\}_0^{k-1}] \\
&= \hat{x}_k^L - \hat{x}_{k|k-1}^L = K_k^L \tilde{y}_k,
\end{aligned} \tag{4.7}
$$

where $\tilde{y}_k = y_k - C\hat{x}_{k|k-1}^L$ is the measurement innovation.

Lemma 4.1 *The estimate innovation sequence $\{z\}_0^k$ has the following properties:*

1) $z_k \sim \mathcal{N}(\mathbf{0}, P_{k|k-1}^L - P_k^L)$.

2) $\{z\}_0^k$ is an independent sequence.

Lemma 4.2 *$P_{k|k-1}^L - P_k^L \geq 0$, rank$(P_{k|k-1}^L - P_k^L) = q$, where $q =$ rank(C) and $q = n$ if and only if C is full column rank.*

Proof: Since $Q, R > 0$, we have $P_{k|k-1}^L > 0$ and $(CP_{k|k-1}^L C' + R)^{-1} > 0$, and then $(CP_{k|k-1}^L C' + R)^{-1} = L_k' L_k$, where L_k is of full rank. We have

$$
\begin{aligned}
P_{k|k-1}^L - P_k^L &= P_{k|k-1}^L C'(CP_{k|k-1}^L C' + R)^{-1} CP_{k|k-1}^L \\
&= (L_k C P_{k|k-1}^L)' L_k C P_{k|k-1}^L \geq 0.
\end{aligned} \tag{4.8}
$$

Thus, rank$(P_{k|k-1}^L - P_k^L) =$ rank$(C) = q$. $\square$

Since $P_{k|k-1}^L - P_k^L \geq 0$, there exists a unitary matrix $U_k \in \mathbb{R}^{n\times n}$ such that

$$
U_k'(P_{k|k-1}^L - P_k^L)U_k = \text{diag}(\Lambda_k, \mathbf{0}_{n-q}) \tag{4.9}
$$

where $\Lambda_k = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_q)$ and $\lambda_1, \lambda_2, \cdots, \lambda_q$ are the positive eigenvalues of $P_{k|k-1}^L - P_k^L$. Define $F_k = [(\Lambda_k)^{-\frac{1}{2}}, \mathbf{0}_{q\times(n-q)}]U_k'$. Then the normalized estimate innovation vector is

$$
\epsilon_k = F_k z_k. \tag{4.10}
$$

Obviously, $\epsilon_k \sim \mathcal{N}(\mathbf{0}, I_q)$. Without loss of generality, we will use n instead of q for the remainder of this chapter.

For each time k, define the latest transmission before time k

$$t_k = \max\{t \in \mathbb{Z} : t < k, \gamma_t = 1\}. \tag{4.11}$$

Let τ_k and τ_k^+ denote the elapsed time of the latest transmission of before time k and up to and including time k, respectively, i.e.,

$$\tau_k = k - t_k, \qquad \tau_k \in \{1, 2, \ldots, k\} \tag{4.12}$$

$$\tau_k^+ = k - \max\{t \in \mathbb{Z} : t \leq k, \gamma_t = 1\}, \quad \tau_k^+ \in \{0, 1, \ldots, k\}. \tag{4.13}$$

When the time index is clear from the context, we will write t_k as t, τ_k as τ and τ_k^+ as τ^+ for simplicity. Then, it follows from (1.16), (1.18),, (4.7) that $\hat{x}_k^L$ can be represented as

$$\begin{aligned} \hat{x}_k^L &= A^\tau \hat{x}_t^L + A^{\tau-1} z_{t+1} + \cdots + z_k \\ &= A^\tau \hat{x}_t^L + A^{\tau-1} F_{t+1}^\dagger \epsilon_{t+1} + \cdots + F_k^\dagger \epsilon_k \\ &= A^\tau \hat{x}_t^L + A^F \epsilon_{t+1,k} \end{aligned} \tag{4.14}$$

where $F_k^\dagger = U_k[(\Lambda_k)^{\frac{1}{2}}, \mathbf{0}_{q\times(n-q)}]'$, $A^F = [A^{\tau-1}F_{t+1}^\dagger, \ldots, AF_{k-1}^\dagger, F_k^\dagger]$ and $\epsilon_{t+1,k} = [\epsilon_{t+1}', \epsilon_{t+2}', \ldots, \epsilon_k']'$. Obviously, $A^F\epsilon_{t+1,k}$ contains all innovations in the local estimate at the current time k since the latest transmission.

We propose a novel hybrid data-time-driven communication scheme θ^γ based on the cumulative innovation $\epsilon_{t+1,k}$, the time-varying threshold δ_k and the maximum elapsed time M. The corresponding decision variable γ_k is defined as

$$\gamma_k = \begin{cases} 1 & \text{if } ||\epsilon_{t+1,k}||^2 > \delta_k \text{ or } \tau > M \\ 0 & \text{if } ||\epsilon_{t+1,k}||^2 \leq \delta_k \text{ and } \tau \leq M \end{cases} \tag{4.15}$$

where $||\epsilon_{t+1,k}||^2$ is a cumulative innovation statistic and $\delta_k \in (0, \delta_{max})$ is a time-varying threshold, which can be pre-specified or designed based on a specified mathematical model of optimization. If $\gamma_k = 1$, then the sensor sends $\mathbb{E}[x_k|\{y\}_0^k] - \mathbb{E}[x_k|\{y\}_0^t] = \hat{x}_k^L - A^\tau \hat{x}_t^L = A^F \epsilon_{t+1,k} \triangleq \tilde{\epsilon}_{k,\tau}$, which is the cumulative estimate innovation at time k since the latest transmission. Information "no data transmission" at time k, corresponding to $||\epsilon_{t+1,k}||^2 \leq \delta_k$, can be and should be used to improve estimation performance. Note that $||\epsilon_{t+1,k}||^2 = ||\epsilon_{t+1}||^2 + ||\epsilon_{t+2}||^2 + \cdots + ||\epsilon_k||^2$.

Lemma 4.3 *Under our communication scheme, the probability of transmission depends on the value of τ_{k-1}^+ and $(\delta_{k-j}, \cdots, \delta_k)$ and is given by*

$$\mathbb{P}(\gamma_k = 1|\tau_{k-1}^+ = j) \tag{4.16}$$
$$= 1 - \begin{cases} \mathbb{P}(\chi_n^2 \leq \delta_k) & \textit{if } j = 0 \\ \dfrac{\int\limits_{D_{k-j}} f(u_j;n)\mathrm{d}u_j \cdots \int\limits_{D_{k-1}} Pr(\chi_n^2 \leq \delta_k - \sum_{l=1}^{j} u_l) f(u_1;n)\mathrm{d}u_1}{\int\limits_{D_{k-j}} f(u_j;n)\mathrm{d}u_j \cdots \int\limits_{D_{k-2}} Pr(\chi_n^2 \leq \delta_{k-1} - \sum_{l=2}^{j} u_l) f(u_2;n)\mathrm{d}u_2} & \textit{if } j = 1, 2, \cdots, M-1 \\ 0 & \textit{if } j = M \end{cases}$$

where $D_{k-l} = [0, \delta_{k-l} - \sum_{h=l+1}^{j} u_h]$, $f(u_l; n)$ is the pdf of χ_n^2 ($l \in \{1, 2, \cdots, j\}$).

Proof: For $j = 1, 2, \ldots, M-1$,

$$\mathbb{P}(\gamma_k = 0|\tau_{k-1}^+ = j) \tag{4.17}$$
$$= \mathbb{P}\left(\sum_{h=k-j}^{k} ||\epsilon_h||^2 \leq \delta_k \,\middle|\, \sum_{h=k-j}^{l} ||\epsilon_h||^2 \leq \delta_l, l \in \{k-j, \cdots, k-1\} \right) \tag{4.18}$$
$$= \frac{\mathbb{P}(||\epsilon_l||^2 \leq (\delta_l - \sum_{h=k-j}^{l-1} ||\epsilon_h||^2), l \in \{k-j, \cdots, k\})}{\mathbb{P}(||\epsilon_l||^2 \leq (\delta_l - \sum_{h=k-j}^{l-1} ||\epsilon_h||^2), l \in \{k-j, \cdots, k-1\})} \tag{4.19}$$
$$= \frac{\int\cdots\int\limits_{D_{k-j}\times\cdots D_k} f_{||\epsilon_{k-j}||^2, \cdots, ||\epsilon_k||^2}(u_j, \cdots, u_0)\mathrm{d}u_0 \cdots \mathrm{d}u_j}{\int\cdots\int\limits_{D_{k-j}\times\cdots D_{k-1}} f_{||\epsilon_{k-j}||^2, \cdots, ||\epsilon_{k-1}||^2}(u_j, \cdots, u_1)\mathrm{d}u_1 \cdots \mathrm{d}u_j} \tag{4.20}$$
$$= \frac{\int\limits_{D_{k-j}} f(u_j;n)\mathrm{d}u_j \cdots \int\limits_{D_{k-1}} Pr(\chi_n^2 \leq \delta_k - \sum_{l=1}^{j} u_l) f(u_1;n)\mathrm{d}u_1}{\int\limits_{D_{k-j}} f(u_j;n)\mathrm{d}u_j \cdots \int\limits_{D_{k-2}} Pr(\chi_n^2 \leq \delta_{k-1} - \sum_{l=2}^{j} u_l) f(u_2;n)\mathrm{d}u_2}. \tag{4.21}$$

The cases of $j = 0$ and $j = M$ follows directly from the definition of γ_k in (4.15). $\square$

4.4 The remote MMSE estimator

This section shows how a remote estimator uses the sensor data to obtain the optimal state estimate $\hat{x}_k$ in (4.4) and its MSE matrix P_k in (4.5) analytically under our communication scheme. We will also give an upper bound of the trace of P_k.

Under our communication scheme, $\mathcal{I}_k = \{1, \hat{x}_k^L - A^\tau \hat{x}_t^L\}$ or $\{0\}$. Here $\{1\}$ or $\{0\}$ indicates $\gamma_k = 1$ or 0. Let the region $\Omega_k = \{\epsilon_{t+1,k} \in \mathbb{R}^{\tau n} : ||\epsilon_{t+1,k}||^2 \leq \delta_k\}$ carry information about the sensor data with $\gamma_k = 0$. Note that $\gamma_k = 0$ implies $\gamma_i = 0$ $(i \in \{t+1, \cdots, k-1, k\})$. Thus, if $\gamma_k = 0$, then $\{\mathcal{I}\}_0^k = \{\{\mathcal{I}\}_0^t, \Omega_{t+1}, \ldots, \Omega_{k-1}, \Omega_k\}$.

Lemma 4.4

$$\mathbb{E}[\epsilon_{t+1,k} | \{\mathcal{I}\}_0^t, \Omega_{t+1}, \ldots, \Omega_{k-1}, \Omega_k] = 0 \tag{4.22}$$

$$\mathbb{E}[\epsilon_{t+1,k}(\epsilon_{t+1,k})' | \{\mathcal{I}\}_0^t, \Omega_{t+1}, \ldots, \Omega_{k-1}, \Omega_k] = \eta(\delta_{t+1}^k) \tag{4.23}$$

$$= \mathit{diag}(\eta_{t+1}(\delta_{t+1}^k), \eta_{t+1}(\delta_{t+1}^k), \cdots, \eta_k(\delta_{t+1}^k)) \tag{4.24}$$

where $\delta_{t+1}^k = (\delta_{t+1}, \cdots, \delta_k)$,
$\eta_i(\delta_{t+1}^k) = \frac{1}{n} \times \dfrac{\int_{D_{t+1}} f(u_{t+1};n)\mathrm{d}u_{t+1} \cdots \int_{D_i} u_i f(u_i;n)\mathrm{d}u_i \cdots \int_{D_k} f(u_k;n)\mathrm{d}u_k}{\int_{D_{t+1}} f(u_{t+1};n)\mathrm{d}u_{t+1} \cdots \int_{D_{k-1}} Pr(\chi_n^2 \leq \delta_k - \sum_{h=t+1}^{k-1} u_h) f(u_{k-1};n)\mathrm{d}u_{k-1}} I_n$, $D_l = [0, \delta_l - \sum_{h=t+1}^{l-1} u_h]$, $f(u_l; n)$ *is the pdf of* χ_n^2 *(*$i, l \in \{t+1, \cdots, k\}$*).*

Proof: Unconditionally, $\epsilon_k \sim \mathcal{N}(\mathbf{0}, I_n)$ according to Lemma 4.1 and (4.10). Also, $\epsilon_{t+1}, \epsilon_{t+2}, \ldots, \epsilon_k$ are independent. Thus, unconditionally, $\epsilon_{t+1,k} \sim \mathcal{N}(\mathbf{0}, I_{(k-t)n})$.

It is clear that $\epsilon_{t+1,k}$ and $\{\mathcal{I}\}_0^t$ are independent. Since the pdf of $\epsilon_{t+1,k}$ and the integration region, i.e., the common part of regions $\Omega_{t+1}, \cdots, \Omega_{k-1}, \Omega_k$ are all symmetrical about the origin, $\mathbb{E}[\epsilon_{t+1,k} | \{\mathcal{I}\}_0^t, \Omega_{t+1}, \cdots, \Omega_{k-1}, \Omega_k] = \mathbb{E}[\epsilon_{t+1,k} | \Omega_{t+1}, \cdots, \Omega_{k-1}, \Omega_k] = 0$.

$$\mathbb{E}[\epsilon_{t+1,k}(\epsilon_{t+1,k})^{'}|\{\mathcal{I}\}_0^t,\Omega_{t+1},\cdots,\Omega_{k-1},\Omega_k] \tag{4.25}$$

$$= \left[\begin{array}{ccc|c} \epsilon_{t+1}(\epsilon_{t+1})^{'} & \ldots & \epsilon_{t+1}(\epsilon_k)^{'} & \\ \vdots & \ddots & \vdots & \Omega_{t+1},\cdots,\Omega_{k-1},\Omega_k \\ \epsilon_k(\epsilon_{t+1})^{'} & \ldots & \epsilon_k(\epsilon_k)^{'} & \end{array}\right]. \tag{4.26}$$

For $\forall i,j \in \{t+1,\cdots,k-1,k\}$, if $i \neq j$, then $\mathbb{E}[\epsilon_i(\epsilon_j)^{'}|\Omega_{t+1},\cdots,\Omega_{k-1},\Omega_k] = 0$ since 1) ϵ_i and ϵ_j are both odd functions; 2) the pdf of ϵ_i, ϵ_j and the integration region are symmetric around the origin.

If $i = j$,

$$\mathbb{E}[\epsilon_i(\epsilon_i)^{'}|\Omega_{t+1},\cdots,\Omega_{k-1},\Omega_k] \tag{4.27}$$

$$= \frac{1}{n}\mathbb{E}[(\epsilon_i)^{'}\epsilon_i|\Omega_{t+1},\cdots,\Omega_{k-1},\Omega_k]I_n \tag{4.28}$$

$$= \frac{1}{n}\mathbb{E}\left[||\epsilon_i||^2 \middle| \sum_{h=t+1}^{l}||\epsilon_h||^2 \leq \delta_l, l \in \{t+1,\cdots,k-1,k\}\right]I_n \tag{4.29}$$

$$= \frac{1}{n} \times \frac{\int\limits_{D_{t+1}\times\cdots D_k}\cdots\int u_i f_{||\epsilon_{t+1}||^2,\cdots,||\epsilon_k||^2}(u_{t+1},\cdots,u_k)\mathrm{d}u_k\cdots\mathrm{d}u_{t+1}}{n\int\limits_{D_{t+1}\times\cdots D_k}\cdots\int f_{||\epsilon_{t+1}||^2,\cdots,||\epsilon_k||^2}(u_{t+1},\cdots,u_k)\mathrm{d}u_k\cdots\mathrm{d}u_{t+1}}I_n \tag{4.30}$$

$$= \frac{1}{n} \times \frac{\int\limits_{D_{t+1}} f(u_{t+1};n)\mathrm{d}u_{t+1}\cdots\int\limits_{D_i} u_i f(u_i;n)\mathrm{d}u_i\cdots\int\limits_{D_k} f(u_k;n)\mathrm{d}u_k}{\int\limits_{D_{t+1}} f(u_{t+1};n)\mathrm{d}u_{t+1}\cdots\int\limits_{D_{k-1}}\Pr(\chi_n^2 \leq \delta_k - \sum_{h=t+1}^{k-1}u_h)f(u_{k-1};n)\mathrm{d}u_{k-1}}I_n. \tag{4.31}$$

This completes the proof. □

Theorem 4.1 *Consider the remote state estimation with our communication scheme, the MMSE estimator of x_k given $\{\mathcal{I}\}_0^k$ under Assumption 4.1 is*

$$\hat{x}_k = A\hat{x}_{k-1} + \gamma_k(\hat{x}_k^L - A^\tau\hat{x}_t^L), \tag{4.32}$$

$$P_k = P_k^L + (1-\gamma_k)\sum_{i=0}^{\tau-1}\eta_{k-i}(\delta_{t+1}^k)A^i(P_{k-i|k-i-1}^L - P_{k-i}^L)(A^i)'. \tag{4.33}$$

Proof: The proof is easy to obtain based on Theorem 3.1 and Lemma 4.4. □

The computation of P_k involves numerical integration, which is often intractable in general. We try to derive an upper bound for the trace of P_k. First of all, the following definitions and lemmas are given.

Let random vectors $\Xi_1, \Xi_2, \cdots, \Xi_N$ be i.i.d., and $\Xi_j \sim \mathcal{N}(\mathbf{0}, I_n)$ ($j \in \{1, 2, \cdots, N\}$, $N, n \in \mathbb{N}$). Denote the region $d_j = \{[\xi_1', \xi_2', \cdots, \xi_N'] \in \mathbb{R}^{Nn} : \sum_{i=1}^{j} ||\xi_i||^2 \leq \delta_j\}$ ($\delta_j \geq 0$) and $D = d_1 \cap d_2 \cdots \cap d_N$.

Lemma 4.5 *The boundary of D, denoted by ∂D, has the following properties:*

1) ∂D is almost everywhere smooth.

2) For $\forall \boldsymbol{v} \in \mathbb{R}^n$, where $|\boldsymbol{v}| = 1$, $\exists a \in \mathbb{R}$, s.t. $a\boldsymbol{v} \in \partial D$ and

$$b\boldsymbol{v} = \begin{cases} \in D & \text{if } b \in [0, a] \\ \notin D & \text{otherwise.} \end{cases} \tag{4.34}$$

Proof: 1) Since $\partial d_j = \{\xi_i \in \mathbb{R}^n | \sum_{i=1}^{j} |\xi_i|^2 = \delta_j\}$ are smooth and $D = d_1 \cap d_2 \cdots \cap d_N$, ∂D is smooth almost everywhere.

2) Since d_j are all closed convet sets, any intersection of any family of closed sets is closed and the intersection of any collection of convex sets is convex, we have that D is a closed convex set. Then $\exists a \in \mathbb{R}$, s.t. $a\mathbf{v} \in \partial D$, where $\mathbf{v}$ is an arbitary unit vector in $\mathbb{R}^n$.

Also, since D is a convex set, then

$$b\mathbf{v} = \begin{cases} \in D & \text{if } b \in [0, a] \\ \notin D & \text{otherwise.} \end{cases} \tag{4.35}$$

This completes the proof. □

Lemma 4.6 *The following inequality holds:*

$$E\left[\sum_{i=1}^{N} ||\Xi_i||^2 \middle| d_1, d_2, \cdots, d_N\right] \leq E\left[\sum_{i=1}^{N} ||\Xi_i||^2 \middle| d_N\right] \tag{4.36}$$

where the equality holds if and only if $\delta_j \geq \delta_N$ *for* $\forall j \in \{1, 2, \cdots, N-1\}$.

Proof: $[\Xi_1', \Xi_2', \cdots, \Xi_N'] \sim \mathcal{N}(\mathbf{0}, I_{Nn})$ since $\Xi_j \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mathbf{0}, I_n)$. Now we switch to the spherical coordinates.

$$f_{\Xi_1,\Xi_2,\cdots,\Xi_N}(\xi_1, \xi_2, \cdots, \xi_N)\mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N \tag{4.37}$$

$$= f(r, \Phi) r^{Nn-1} \mathrm{d}r\mathrm{d}\Phi = \frac{1}{U} f_r(r) r^{Nn-1} \mathrm{d}r\mathrm{d}\Phi \tag{4.38}$$

where r is the radius, Φ is the angle, and U is the surface area of the Nn-dimensional ball. The second equality holds because the radius and the angle are independent for a Gaussian random vector.

Define $g(r) = r^2$ and $D = d_1 \cap d_2 \cdots \cap d_N$. Then

$$E\left[\sum_{i=1}^{N} ||\Xi_i||^2 \middle| d_1, d_2, \cdots, d_N\right] = E[g(r)|D] = \frac{\int_\Omega \frac{1}{U} \int_0^{r_\Omega^D} g(r) f_r(r) r^{Nn-1} \mathrm{d}r\mathrm{d}\Phi}{\int_\Omega \frac{1}{U} \int_0^{r_\Omega^D} f_r(r) r^{Nn-1} \mathrm{d}r\mathrm{d}\Phi} \tag{4.39}$$

where r_Ω^D is the distance from the origin to ∂D at the angle Ω.

Let $P(X) = \int_0^x g(r) f(r) r^{Nn-1} \mathrm{d}r$, $Q(X) = \int_0^x f(r) r^{Nn-1} \mathrm{d}r$, $R(X) = \frac{P(x)}{Q(x)}$. Then

$$\frac{\mathrm{d}R(x)}{\mathrm{d}x} = \frac{g(x) f(x) x^{Nn-1}}{Q(x)} + \frac{-f(x) x^{Nn-1} \int_0^x g(r) f(r) r^{Nn-1} \mathrm{d}r}{Q^2(x)} \tag{4.40}$$

$$= \frac{f(x) x^{Nn-1}}{Q(x)} \Big(g(x) - \frac{\int_0^x g(r) f(r) r^{Nn-1} \mathrm{d}r}{Q(x)} \Big) \tag{4.41}$$

$$\geq \frac{f(x) x^{Nn-1}}{Q(x)} \Big(g(x) - \frac{g(x) \int_0^x f(r) r^{Nn-1} \mathrm{d}r}{Q(x)} \Big) = 0. \tag{4.42}$$

Consequently, $R(x)$ is monotonically nondecreasing with x.

Since $r_\Omega^D \leq \sqrt{\delta_N}$, we have $R(r_\Omega^D) \leq \frac{P(\sqrt{\delta_N})}{Q(\sqrt{\delta_N})}$ for $\forall \Omega$.

$$E[g(r)|D] = \frac{\int_\Omega P(r_\Omega^D) \mathrm{d}\Phi}{\int_\Omega Q(r_\Omega^D) \mathrm{d}\Phi} = \frac{\int_\Omega R(r_\Omega^D) Q(r_\Omega^D) \mathrm{d}\Phi}{\int_\Omega Q(r_\Omega^D) \mathrm{d}\Phi} \tag{4.43}$$

$$\leq \frac{\int_\Omega \frac{P(\sqrt{\delta_N})}{Q(\sqrt{\delta_N})} Q(r_\Omega^D) \mathrm{d}\Phi}{\int_\Omega Q(r_\Omega^D) \mathrm{d}\Phi} = \frac{P(\sqrt{\delta_N})}{Q(\sqrt{\delta_N})} = E[g(r)|d_n]. \tag{4.44}$$

Obviously, the equal sign holds if and only if $D = d_n$, i.e., $\delta_j \geq \delta_N$ for $\forall j \in \{1, 2, \cdots, N-1\}$.
$\square$

Theorem 4.2 *If* $\gamma_k = 0$,

$$tr(P_k) \leq tr(P_k^L) + \tau n \beta(\tau, \delta_k) \lambda_{max}(A_k^P), \tag{4.45}$$

where $\beta(\tau, \delta_k) = \frac{Pr(\chi^2_{\tau n+2} \leq \delta_k)}{Pr(\chi^2_{\tau n} \leq \delta_k)}$, $A_k^P = diag(A^i(P^L_{k-i|k-i-1} - P^L_{k-i})(A^i)')$, $i \in \{\tau - 1, \cdots, 0\}$.

Proof: If $\gamma_k = 0$,

$$\begin{aligned}
\text{tr}(P_k) &= \text{tr}(P_k^L) + \text{tr}\left(\sum_{i=0}^{\tau-1} \eta_i(\delta_{t+1}^k) A^i (P^L_{k-i|k-i-1} - P^L_{k-i})(A^i)' \right) && (4.46)\\
&= \text{tr}(P_k^L) + \text{tr}(\eta(\delta_{t+1}^k) A_k^P) && (4.47)\\
&\leq \text{tr}(P_k^L) + \text{tr}(\eta(\delta_{t+1}^k)) \lambda_{\max}(A_k^P) && (4.48)\\
&= \text{tr}(P_k^L) + \mathbb{E}\left[\sum_{i=t+1}^{k} ||\epsilon_i||^2 \middle| \Omega_{t+1}, \cdots, \Omega_{k-1}, \Omega_k \right] \lambda_{\max}(A_k^P) && (4.49)\\
&\leq \text{tr}(P_k^L) + \mathbb{E}\left[\sum_{i=t+1}^{k} ||\epsilon_i||^2 \middle| \Omega_k \right] \lambda_{\max}(A_k^P) && (4.50)\\
&= \text{tr}(P_k^L) + \tau n \beta(\tau, \delta_k) \lambda_{\max}(A_k^P), && (4.51)
\end{aligned}$$

where the second inequality holds due to Lemma 4.6, and the last equality follows from Lemma 3.3. $\square$

4.5 Reformulation of the optimization problem and implementation

In this section, we show that Problem 4.1 with our communication scheme can be reformulated as an MDP problem under reasonable assumptions. Then an algorithmic approach is given based on the MDP model. Finally, we extend our results to the optimal average cost problem.

Assumption 4.3 *For each time* k, $\delta_{t+1} \geq \delta_{t+2} \cdots \geq \delta_k$.

Assumption 4.4 *$P_{k|k-1}^L = \hat{P}$ and $P_k^L = \bar{P}$, for $\forall\ k \in \mathbb{Z}$, where $\hat{P}$ is the unique positive definite solution of the algebraic Riccati equation*

$$\hat{P} = A(\hat{P} - \hat{P}C'(C\hat{P}C' + R)^{-1}C\hat{P})A' + Q, \tag{4.52}$$

and $\bar{P}$ is given by

$$\bar{P} = \hat{P} - \hat{P}C'(C\hat{P}C' + R)^{-1}C\hat{P}. \tag{4.53}$$

Remark 4.1 *Assumption 4.3 comes into play if an intuition applies: the more data packet drops accumulate consecutively, the lower the threshold should be to facilitate sending of a data packet at the next time instant.*

Remark 4.2 *Since the pair (A,C) is observable, the prior state covariance $P_{k|k-1}^L$ and the posterior covariance P_k^L converge to the steady-state values $\hat{P}$ and $\bar{P}$, respectively, exponentially fast. Meanwhile, we consider an infinite horizon, and we omit the transient estimation process at the sensor side. As a result, we use Assumption 4.4 throughout the optimization problem.*

Under Assumptions 4.3, $\epsilon_{t+1,k} \in \Omega_k \Rightarrow \epsilon_{t+1,k-1} \in \Omega_{k-1}, \cdots, \epsilon_{t+1,t+1} \in \Omega_{t+1}$, i.e., $\epsilon_{t+1,i} \in \Omega_i$ is sufficient but not necessary for $\epsilon_{t+1,j} \in \Omega_j (i \geq j \geq t+1)$ since $||\epsilon_{t+1,i}||^2 \geq ||\epsilon_{t+1,j}||^2$. Thus, if $\gamma_k = 0$, then $\{\mathcal{I}\}_0^k = \{\{\mathcal{I}\}_0^t, \Omega_{t+1}, \ldots, \Omega_{k-1}, \Omega_k\} = \{\{\mathcal{I}\}_0^t, \Omega_k\}$ since $\epsilon_{t+1,k} \in \Omega_k$ implies $\epsilon_{t+1,k-1} \in \Omega_{k-1}, \cdots, \epsilon_{t+1,t+1} \in \Omega_{t+1}$.

As a consequence, under Assumptions 4.3 and 4.4, according to Lemma 3.3 and (4.13), P_k in (4.33) can be rewritten as

$$P_k = \bar{P} + \beta(\tau_k^+, \delta_k) \sum_{i=0}^{\tau_k^+ - 1} A^i(\hat{P} - \bar{P})(A^i)' \tag{4.54}$$

where $\beta(0, \delta_k) = 0$ and $\sum_{i=0}^{-1}(\cdot) = 0$.

4.5.1 Stochastic DP model (S, U, W)

We now reformulate Problem 4.1. We consider a discrete-time dynamical system. For all $k \in \{1, 2, \cdots\}$, the system state is defined as $s_k = (\tau_{k-1}^+, \delta_{k-1}) \in S$, where $S = \{0, 1, \cdots, M\} \otimes (0, \delta_{max})$ is the state space; the system control is defined as $u_k = \mu_k(s_k) = \mu_k(\tau_{k-1}^+, \delta_{k-1}) \in U$, where $U = (0, \delta_{max})$ is the control space; the disturbance is defined as $w_k = \mathbb{1}(||\epsilon_{k-\tau_{k-1}^+, k}||^2 > u_k) \in W$, where $W = \{0, 1\}$ is the disturbance space. Under Assumption 4.3, w_k are characterized by probabilities $\Pr(\cdot|s_k, u_k)$ defined on W, where

$$\Pr(w_k = 0|(\tau_{k-1}^+ = \tau^+, \delta_{k-1} = \delta), u_k = u) = \begin{cases} \Pr(\chi_n^2 \leq u) & \text{if } \tau^+ = 0 \\ \frac{\Pr(\chi^2_{(\tau^+ + 1)n} \leq u)}{\Pr(\chi^2_{\tau^+ n} \leq \delta)} & \text{if } \tau^+ > 0 \end{cases} \tag{4.55}$$

$$\Pr(w_k = 1|(\tau_{k-1}^+ = \tau^+, \delta_{k-1} = \delta), u_k = u) = 1 - \Pr(w_k = 0|(\tau_{k-1}^+ = \tau^+, \delta_{k-1} = \delta), u_k = u). \tag{4.56}$$

Obviously, the state s_k evolves according to the following stationary discrete-time dynamical system

$$s_{k+1} = f(s_k, u_k, w_k), \qquad k = 1, 2, \cdots \tag{4.57}$$

where $f : S \times U \times W \mapsto S$ and is given by

$$\begin{cases} \tau_k^+ = (\tau_{k-1}^+ + 1)(1 - w_k) \\ \delta_k = u_k. \end{cases} \tag{4.58}$$

We further define the admissible policy at time k as $\mu_k = \{\mu \in S \mapsto U | \mu(\tau^+, \delta) \in U,\ \mu(\tau^+, \delta) \leq \delta \text{ for } \tau^+ > 0\}$ and the cost per stage $g : S \times U \times W \mapsto \mathbb{R}$ as

$$g(s_k, u_k, w_k) = \text{tr}(\bar{P}) + \beta(\tau_{k-1}^+ + 1, u_k)\text{tr}\left(\sum_{i=0}^{\tau_{k-1}^+} A^i(\hat{P} - \bar{P})(A^i)'\right)(1 - w_k) + \kappa w_k. \tag{4.59}$$

According to Theorem 3.2, the cost per stage g is bounded.

Given an initial state $s_1 = (\tau_0^+, \delta_0)$, we wish to find an admissible policy $\pi = \{\mu_1, \cdots, \mu_k, \cdots\}$ that minimizes the cost function

$$J_\pi(s_1) = \lim_{N\to\infty} \underset{\substack{w_k \\ k=1,2,\cdots}}{E} \left[\sum_{k=1}^{N} \alpha^{(k-1)} g(s_k, \mu_k(s_k), w_k) \right] \tag{4.60}$$

subject to the system (4.57), where $0 \le \alpha < 1$ is the discount factor.

Let Π be the set of admissible policies π. Then the optimal cost function J^* is defined by

$$J^*(s) = \min_{\pi\in\Pi} J_\pi(s), \qquad s \in S. \tag{4.61}$$

Note that if assumption 4.3 and 4.4 are satisfied, the differences between Problem 4.1 under our communication scheme and the optimization problem (4.61) arise from the construct that the policy is Markovian and deterministic (MD), in contrast to the general decision rule which is history dependent and randomized (HR). It will be explained in Remark 4.3 that the optimal cost under MD construct is equal to the one that uses HR.

Let

$$(TJ)(s) = \min_{u\in U(s)} \underset{\substack{w_k \\ k=1,2,\cdots}}{E} [g(s,u,w) + \alpha J(f(s,u,w))], \qquad s \in S. \tag{4.62}$$

Then the optimal cost function J^* in (4.61) satisfies Bellman's equation [116]

$$J^*(s) = (TJ^*)(s). \tag{4.63}$$

Furthermore, J^* is the unique solution of (4.63) with the bounded function g.

4.5.2 Finite-state MDP model (S, U, p, g) and algorithmic design

We further approximate the state space and control space over finite lattices, which enables us to formulate the optimization problem (4.61) as a finite-state MDP problem (S, U, p, g), where

(1) State space $S = \{(\tau^+, \delta) | (\tau^+, \delta) \in \{0, 1, \cdots, M\} \otimes \{\zeta, 2\zeta, \cdots, N\zeta\}\}$, where ζ is the lattice spacing and $N = \lceil \delta_{max}/\zeta \rceil$;

(2) Control space $U = \{\zeta, 2\zeta, \cdots, N\zeta\}\}$;

(3) The transition probability is stationary. Define $s = (\tau^+_{k-1}, \delta_{k-1})$ and $s' = (\tau^+_k, \delta_k)$, according to (4.55), (4.56) and (4.57),

$$
\begin{aligned}
p_{ss'}(u) &= \Pr(\tau^+_k = \tau^{+'}, \delta_k = \delta' | \tau^+_{k-1} = \tau^+, \delta_{k-1} = \delta, u_k = u) \\
&= \delta_{u\delta'} \begin{cases} \Pr(\chi^2_n \le u) & \text{if } \tau^+ = 0, \tau^{+'} = 1 \\ 1 - \Pr(\chi^2_n \le u) & \text{if } \tau^+ = 0, \tau^{+'} = 0 \\ 0 & \text{if } \tau^+ = 0, \tau^{+'} \in \{2, \cdots, M\} \\ \mathbb{1}(u \le \delta) \begin{cases} \frac{\Pr(\chi^2_{(\tau^+ + 1)n} \le u)}{\Pr(\chi^2_{\tau^+ n} \le \delta)} & \text{if } \tau^+ \ne 0, \tau^{+'} = \tau^+ + 1 \\ 1 - \frac{\Pr(\chi^2_{(\tau^+ + 1)n} \le u)}{\Pr(\chi^2_{\tau^+ n} \le \delta)} & \text{if } \tau^+ \ne 0, \tau^{+'} = 0 \\ 0 & \text{if } \tau^+ \ne 0, \tau^{+'} \ne \tau^+ + 1, \tau' \ne 0 \end{cases} \end{cases}
\end{aligned}
\tag{4.64}
$$

(4) The expected cost per stage

$$g(s, u) = \sum_{s'} p_{ss'}(u) \hat{g}(s, u, s') \tag{4.65}$$

where the cost of using u at state s and moving to state s'

$$\hat{g}(s, u, s') = \begin{cases} \mathrm{tr}(\bar{P}) + (\beta(\tau^{+'}, u)\mathrm{tr}(\sum_{i=0}^{\tau^{+'}-1} A^i(\hat{P} - \bar{P})(A^i)') & \text{if } \tau^{+'} > 0 \\ \mathrm{tr}(\bar{P}) + \kappa & \text{if } \tau^{+'} = 0. \end{cases} \tag{4.66}$$

The mapping T of (4.62) can be written as

$$(TJ)(s) = \min_{u \in U(s)} [g(s, u) + \alpha \sum_{s'} p_{ss'}(u) J(s')]. \tag{4.67}$$

Theorem 4.3 *For the infinite horizon discounted cost MDP problem (S, U, p, g), there exists an* *optimal stationary policy μ such that*

$$J_\mu(s) = J^*(s) \qquad \textit{for } \forall s \in S \tag{4.68}$$

where $J^*(s)$ is defined in (4.61).

Proof: The infinite horizon discounted cost MDP problem (S, U, p, g) has a discrete state space S and a finite control space U for each $s \in S$. According to Theorem 6.2.10 in [117], an optimal deterministic stationary policy always exists that attains the optimal cost under Π. Therefore $J_\mu(s)$ must attain $J^*(s)$. $\square$

Remark 4.3 *Let Π^{MD} denote the set of Markovian and deterministic policies and Π^{HR} the set of history dependent and randomized policies. In our MDP framework, we consider Π^{MD}. Theorem 6.2.10 of [117] claims that μ obtains the optimality under most general decision rule Π^{HR}. Since $\mu \in \Pi^{MD}$ and $\Pi^{MD} \subset \Pi^{HR}$, we have $J_\mu(s) = \underset{\pi \in \Pi^{MD}}{\mathit{min}} J_\pi(s) = \underset{\pi \in \Pi^{HR}}{\mathit{min}} J_\pi(s) = J^*(s)$, that is, the optimal cost under our MD construct is equal to the one that uses HR.*

This theorem enable us to content ourselves with the search for an optimal stationary policy. We rewrite our goal in (4.61) as

$$J^*(s) = \min_\mu J_\mu(s). \tag{4.69}$$

We now introduce a theorem that serves as a foundation for our iterative algorithmic design.

Theorem 4.4 *1) The optimal cost function J^* satisfies*

$$J^*(s) = \underset{N \to \infty}{\mathit{lim}} (T^N J)(s) \qquad \textit{for } \forall s \in S. \tag{4.70}$$

2) A stationary policy μ is optimal if and only if $\mu(s)$ attains the minimum in Bellman's equation (4.63) for each $s \in S$; that is,

$$TJ^* = T_\mu J^* \tag{4.71}$$

where $(T_\mu J)(s) = g(s,\mu(s)) + \alpha \sum_{s'} p_{ss'}(\mu(s))J(s')]$.

Proof: Part 1 is a direct result of Proposition 2.1 in [116] since $J: \ S \mapsto \mathbb{R}$ is bounded. Part 2 follows from Proposition 2.3 in [116]. $\square$

Consequently, we can use value iteration in Table 4.1 to get the optimal cost function and policy numerically.

Table 4.1: Value iteration algorithm

Algorithm: Iterative procedure to compute the optimal cost and policy tuple (J,μ)

$i = 0$

$$J^{(0)} = \begin{bmatrix} J^{(0)}(0,\zeta), & J^{(0)}(0,2\zeta), & \cdots & J^{(0)}(0,N\zeta), \\ J^{(0)}(1,\zeta), & J^{(0)}(1,2\zeta), & \cdots & J^{(0)}(1,N\zeta), \\ & & \vdots & \\ J^{(0)}(M,\zeta), & J^{(0)}(M,2\zeta), & \cdots & J^{(0)}(M,N\zeta), \end{bmatrix}$$

repeat
    $\tau^+ = 0$
    repeat
        $x = 1$
        repeat
            $[J^{(i+1)}(s), \mu(s)] = \min_{l} \sum_{\tau^{+'}} \sum_{y} p_{ss'}(l\zeta)[\hat{g}(s, l\zeta, s') + \alpha J^{(i)}(s')]$
            where $s = (\tau^+, x\zeta)$ and $s' = (\tau^{+'}, y\zeta)$
        $x \leftarrow x + 1$
        until $x > N$
    $\tau^+ \leftarrow \tau^+ + 1$
    until $\tau^+ > M$
$i \leftarrow i + 1$
until convergence
return $(J^{(i)}, \mu)$

Definition 4.1 *(Definiton 5.2.1 in [118]) State s' is accessible from state s if there exists a stationary policy μ and an integer k such that*

$$Pr(s_k = s' | s_0 = s, \mu) > 0. \tag{4.72}$$

Our MDP framework invites the following property of degeneracy.

Proposition 4.1 *Given any states $(0, \delta)$, $\forall \delta \in \{\zeta, 2\zeta, \cdots, N\zeta\}$ and an optimal stationary policy μ^*, then:*

1) The state $(0, \delta)$ degenerates w.r.t μ^*, i.e., $\mu^*(0, \delta) = \mu^*(\tau^+ = 0)$, which is independent of the value of δ.

2) If (τ^+, δ_1), (τ^+, δ_2) are both accessible from $(0, \delta)$ under μ^* for $\forall \tau^+ > 0$, then $\delta_1 = \delta_2$.

Proof: Part 1 follows from the fact that the optimality equations for $J(\tau^+ = 0, \delta)$ have identical form for all δ's and, perforce, degenerate all $(\tau^+ = 0, \delta)$ for all δ's.

Part 2 comes from the observation that since any current state $(\tau^+ = 0)$ or $(\tau^+ > 0, \delta_{\tau^+})$ can only transition to $(\tau^+ = 0)$ or $(\tau^+ + 1, \delta_{\tau^+ + 1})$, where $\delta_{\tau^+ + 1}$ is determined by μ^* and current state, with nonzero probability, any accessible state $(\tau^+, \delta_{\tau^+})$ can trace back to $(\tau^+ = 0)$ via a single path. $\square$

This property essentially tells that under our MDP settings, the optimal stationary policy μ^*, given any initial states $(\tau^+ = 0, \delta)$, confines the entire $(M+1) \times N$-dimensional state space to a τ^+-indexed subspace of $(M+1)$-dimension, necessitating the second state parameter δ to collapse to a fixed δ_{τ^+} for the same value of τ^+.

4.5.3 Expected average cost over infinite horizon

We now consider the optimization problem of the expected average cost over the infinite horizon

$$\min_{\pi^\gamma} \lim_{T \to \infty} \frac{1}{T} E^{\pi^\gamma} \left[\sum_{k=1}^{T} (\mathrm{tr}(P_k) + \kappa \gamma_k) \right]. \tag{4.73}$$

Following the lines of subsections A and B in section IV, one could model problem (4.73) also in an MDP framework. It is desired to find a policy $\pi = \{\mu_1, \cdots, \mu_k, \cdots\}$ to minimize the cost function

$$H_\pi(s_1) = \lim_{T \to \infty} \frac{1}{T} E \left[\sum_{k=1}^{N} g(s_k, \mu_k(s_k)) \right]. \tag{4.74}$$

Let Π be the set of admissible policies π. Then the optimal cost function H^* is defined by

$$H^*(S) = \min_{\pi \in \Pi} H_\pi(S). \tag{4.75}$$

It is noteworthy that for arbitrary two states $s = (\tau^+, \delta)$ and $s' = (\tau^{+'}, \delta')$ in the dynamical system (4.57), the accessibility condition holds. This observation helps deliver a conclusion of both theoretical and practical implications (Proposition 2.6 in [116]): the optimal average cost per stage has the same value for arbitrary initial states if the accessibility condition holds and further,

$$\lambda = \lim_{\alpha \to 1} (1-\alpha) J_\alpha^*(s) \qquad \forall s \in S \tag{4.76}$$

where λ is the optimal average cost per stage and J_α^* is J^* in (4.68). That is, $\lambda = H^*(s)$, which is independent of the initial state s.

From a practitioner's point of view, one could, at least in principle, obtain the optimal average cost per stage and the corresponding optimal policy by taking advantage of the converging sequence of $(1-\alpha)J_\alpha^*$. It is the pathway we follow in our quest for the optimal average cost.

4.6 Illustrative example

This section presents an example to illustrate the results in this chapter.

Consider the following stochastic linear system

$$x_{k+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_k + w_k \tag{4.77}$$

$$y_k = \begin{bmatrix} 1 & 1 \\ 0 & 1.3 \end{bmatrix} x_k + v_k \tag{4.78}$$

with $Q = 5I_2$, $R = 2I_2$, $P_0 = 0.3I_2$, $x_0 = [1, 1]'$.

Given the discount factor $\alpha = 0.999$, the maximum elapsed time $M = 6$ and the lattice spacing $\zeta = 0.1$, by applying our MDP approach, the optimal discounted cost J^*, the optimal stationary policy μ^* and the approximated optimal average cost per stage $\hat{\lambda} = (1-\alpha)J^*$, versus different weighted unit transmission costs $\kappa = (5,\ 20,\ 35)$, are obtained and illustrated in Table 4.2. J^* and H^* grow with an increase in the unit transmission cost κ. In addition, for the same value of τ^+, the threshold rises with growing κ. The intuition behind the upshift is that a high expense of communication suppresses the rate of transmission. It harkens back to the goal of our policy that optimally balances communication consumption with estimation quality. Fig. 4.1 depicts the optimal communication rates under different κ (1 to 40) by our MDP approach. The communication rates decrease with κ. It also illustrates the effect of our policy.

Table 4.2: Optimization Results

κ	J^*	U^*							H^*
		$\tau^+ = 0$	$\tau^+ = 1$	$\tau^+ = 2$	$\tau^+ = 3$	$\tau^+ = 4$	$\tau^+ = 5$	$\tau^+ = 6$	
5	4152.8783	0.9	0.6	0.4	0.3	0.2	0.1	0.1	7.0679
20	12836.4439	2.7	2	1.3	0.9	0.6	0.4	0.1	15.7515
35	19607.0952	4.1	3	2	1.3	0.9	0.7	0.1	22.5221

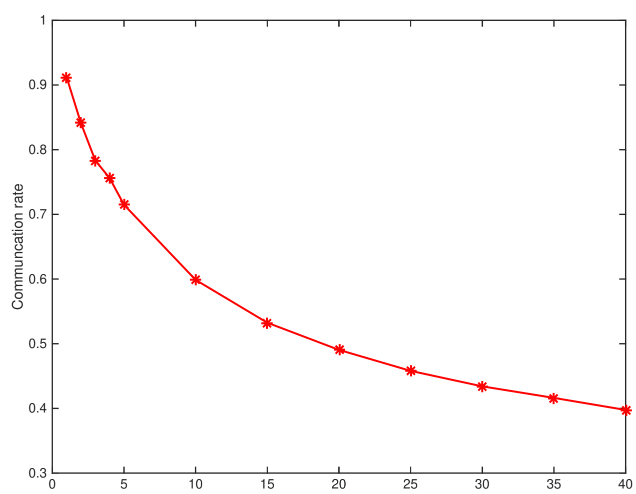


Figure 4.1: Communication rate vs. κ

Define the time averaged sample mean cost as

$$\frac{1}{T}\sum_{k=1}^{T}\overline{(\text{tr}(P_k)+\kappa\gamma_k)} \tag{4.79}$$

where T is total time, $\overline{\text{tr}(P_k)+\kappa\gamma_k}$ is the sample mean of cost per stage obtained from 500 Monte Carlo runs. Fig. 4.2 compares (4.79) with $\hat{\lambda}$ for $\kappa=(5,\ 20,\ 35)$ and shows that they all converge to $\hat{\lambda}$ fast with an increase in runtime. The validity of the approximated theoretical value is thereby manifested.

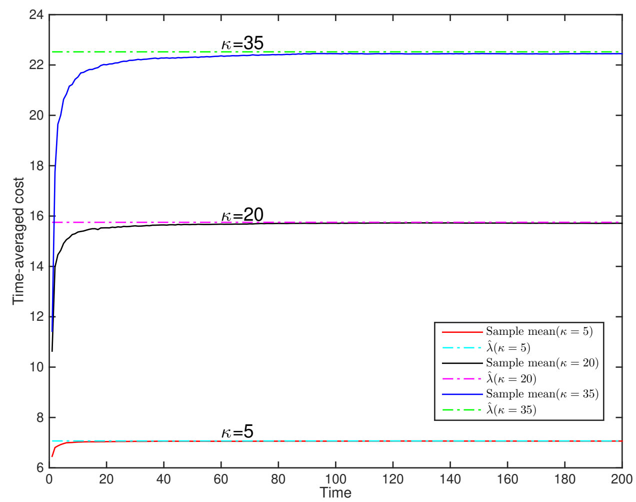


Figure 4.2: Time averaged sample mean cost and $\hat{\lambda}$ for $\kappa=(5,\ 20,\ 35)$

For $\kappa=(5,\ 20,\ 35)$, Fig. 4.3 and Fig. 4.4 depict J^* as a function of M and ζ, respectively. Fig. 4.3 shows that for smaller M, J^* drops. But it converges even at moderately large M. Therefore, the value 6 we used before already works with great precision. It is observed that the trend plateaus faster for a smaller κ, which corroborates the fact that our optimal policy for a smaller k tends to suppress the probability of large τ^+ and therefore the influence of M on J^* is marginal. This effect is also clear from Table 4.2—at $\kappa=5$, U^* for $\tau^+=5$ already drops to 0.1 (the smallest possible threshold). Fig. 4.4 shows the influence of lattice spacing—J^* drops for smaller ζ; yet it converges at least at $\zeta=0.5$. This observation enables us to use the value 0.1 with high confidence. One should notice that the effect of varying ζ is more pronounced for a smaller

κ. That is because smaller κ corresponds to a lower threshold (shown in Table 4.2) for the same τ^+, thereby sensitizing the choice of ζ.

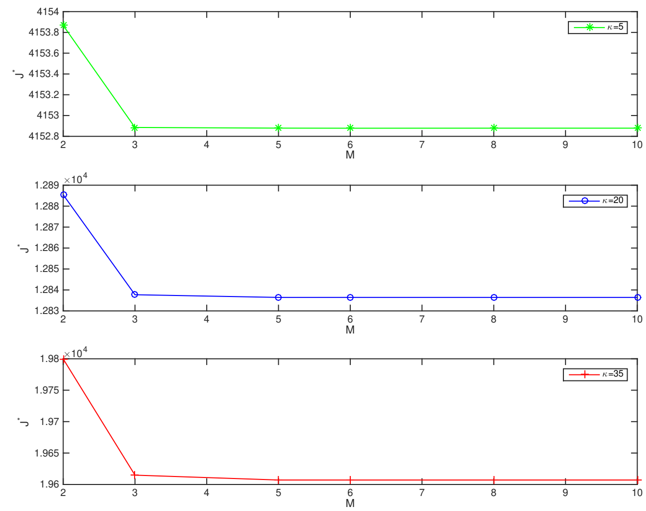


Figure 4.3: J^* for $\kappa = (5,\ 20,\ 35)$ vs. M

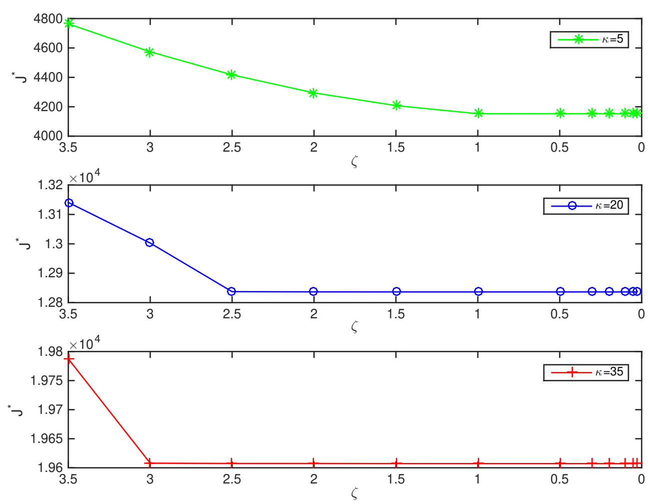
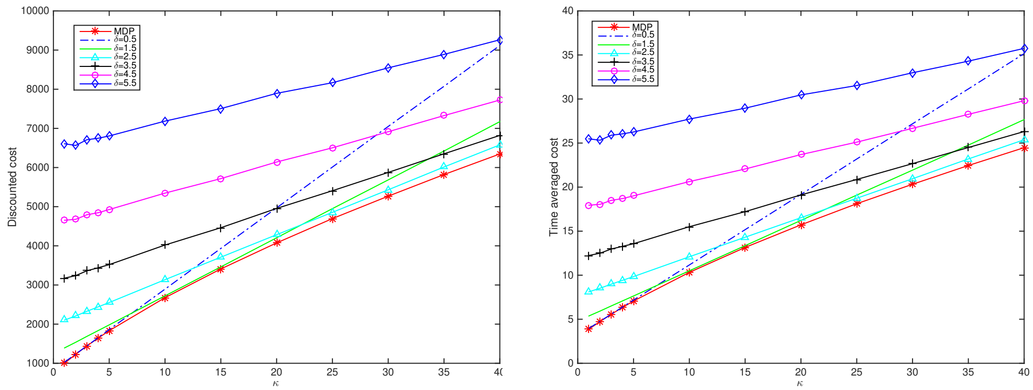


Figure 4.4: J^* for $\kappa = (5,\ 20,\ 35)$ vs. ζ

For our MDP approach and six other fixed thresholding policies with $\delta = 0.5 : 1 : 5.5$, we compare their discounted costs and time averaged sample mean costs for different κ values over 400 time steps and 500 Monte Carlo runs in Fig. 4.5 (a) and (b), respectively. They both confirm



(a) Discounted costs vs. κ

(b) Time averaged sample mean costs vs. κ

Figure 4.5: Discounted costs and time averaged sample mean costs for MDP policy and six fixed thresholding policies with $\delta = 0.5 : 1 : 5.5$ vs. κ

that our optimal policy produces the minimal costs for different κ. Some fixed thresholding policies exhibit close behaviors at some small regions of κ, but it is our policy that consistently yields the best result throughout the entire spectrum of κ. The superiority of our MDP approach clearly stands out in that our MDP methodology builds on solid theoretical grounds that ensures the optimality of the solution through a systematic procedure.

4.7 Summary and conclusions

This chapter focuses on the optimal of a data-driven communication scheme and an estimator to quantitatively balance system performance and limited communication resource. We propose a hybrid data-time-driven communication scheme based on the cumulative estimate innovation, time-varying thresholds and the maximum elapsed time and derive the MMSE estimator accordingly. We present a workable tight upper bound for the estimation error matrix. We define the optimization criterion based on the expected total discounted cost over the infinite horizon and show that our system is of the nature of an MDP, enabling an iterative algorithm to find the optimal cost and optimal policy. We also extend our results to solving the optimal average cost problem. We prove that our framework satisfies DOF reduction; specifically, the current optimal thresholding policy depends on no other state variables except the maximum elapsed time. Our simulation results illustrate the effectiveness of our schemes.

CHAPTER 5

CONCLUSIONS AND OUTLOOK

5.1 Conclusions

The dissertation focuses on the systemic design of proper estimation as well as fusion techniques and data-driven communication schemes to infer the state of a dynamic discrete-time linear system over a wireless network. Our work effectively helps extract and share key information over networks to enhance estimates of system states. Our research are further summarized as follows:

1) Based on one sensor estimation, centralized and distributed estimation fusion architectures, respectively, we jointly construct data-driven communication schemes and optimal estimators/fusers. The proposed data-driven communication schemes can achieve a trade-off between communication costs and estimation performance. The corresponding MMSE estimator/fuser and the MMSE-optimal WLS fuser, optimally using information in the triggering decisions, are proposed to boost estimation performance.

2) We propose new data-driven communication schemes based on cumulative estimate innovation to smartly pick key information and derive the corresponding MMSE estimator/fuser, for both the single-sensor case and the multi-sensor case. First, the cumulative estimate innovation, a new triggering variable, is introduced to reflect the importance of data. It is path-dependent, cumulative and less random. As such, the threshold is more easily determined for a better solution of this mixed estimation-communication problem. The estimator/fuser are optimal under well justified assumptions and have guaranteed stability.

3) We quantitatively calibrate the tradeoff between estimation quality and communication expenses to obtain an optimal data-driven estimator. We have proved, with due mathematical rigor, that the estimation error matrix, conditioned on the information carried by the entire "no data transmission" duration, does not exceed the one conditioned only on the information carried by the last

"no data transmission" timestamp. We further give its workable tight upper bound. Through constructing an auxiliary state vector, the optimal problem incorporating estimation error and consumed resources over an infinite horizon is shown to be representable as an MDP problem. An iterative algorithm is proposed to find the optimal policy and optimal cost. The optimal policy has a peculiar degeneracy property. The computational overhead of the proposed design method is less sensitive to the system dimension compared with that of existing algorithms in the literature.

5.2 Outlook

The following problems need to be further considered and addressed in the future research work, which are potentially the building blocks of more comprehensive approaches to data-driven estimation.

1) Our research so far has been built on ideal channels, forgoing the effects of time delays, packet loss, channel fading, etc. For example, estimators and fusers in our existing framework leverage the general knowledge of non-transmission region should the sensor chooses not to transmit. However, the introduction of packet loss impinges on the efficacy our original design because estimators/fusers cannot distinguish between an event of non-transmission from a sensor and that of packet loss. Packet loss can be modeled as a random process whose statistical properties are either pre-specified or derived from combinations of channels, digital modulation schemes, and transmission power. Our formulation may be modified and augmented to account for packet loss probabilistically. On the other hand, we may attempt to address these network-induced problems from different perspectives by, for instance, taking advantage of potentially simplified modeling of the channel.

2) Estimation fusion brings a number of additional challenges such as correlations between the estimates of different network nodes under distributed fusion architecture. In our original design, however, the coordinations between sensors are not accounted for and the decision processes of different nodes are presumed independent. Under DDC framework, redundancy and interdependency between data from different nodes incur ramifications. In addition, this dissertation only

considers systems with fusion center; however, no fusion center exists for estimation fusion over a peer-to-peer network. Such construct awaits investigation.

3) We formulate the JDE problem as one that minimizes a weighted sum of estimation error and communication cost. We are considering an alternative optimization problem that minimizes energy consumption subject to constraints on the estimation performance.

4) Our work is built for discrete-time linear-Gaussian system, it might be more instrumental to consider continuous-time dynamic system with non-synchronous sensor observations. Also, we believe it is worth the effort looking into non-linear system and system with non-Gaussian noises.

5) We have considered the co-design of estimator/fuser and communications scheme, yet the controller is omitted. Therefore, the holistic prescription is needed to incorporate estimator/fuser, controller and communication.

REFERENCES

[1] J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, "A survey of recent results in networked control systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 138–162, 2007.

[2] C. Lu *et al.*, "Real-time wireless sensor-actuator networks for industrial cyber-physical systems," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1013–1024, 2015.

[3] H. Xu, W. Yu, D. Griffith, and N. Golmie, "A survey on industrial internet of things: A cyber-physical systems perspective," *IEEE Access*, vol. 6, pp. 78 238–78 259, 2018.

[4] X. Li, *Fundamentals of Applied Estimation and Filtering*. Course Notes, 2015.

[5] D. Hall and J. Llinas, *Multisensor data fusion*. CRC press, 2001.

[6] S. Haykin, *Communication systems*. John Wiley & Sons, 2008.

[7] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2001.

[8] D. Simon, *Optimal state estimation: Kalman, H_∞, and nonlinear approaches*. John Wiley & Sons, 2006.

[9] F. LEWIS, "Optimal estimation: With an introduction to stochastic control theory," *New York, Wiley-Interscience*, 1986.

[10] B. Chiang and J. Bohg, *Optimal Estimation*. Course Notes, 2022.

[11] C. F. Gauss, *Theory of the motion of the heavenly bodies moving about the sun in conic sections: a translation of Gauss's "Theoria Motus." with an appendix*. 1857.

[12] M. Liggins II, D. Hall, and J. Llinas, *Handbook of multisensor data fusion: theory and practice*. CRC press, 2017.

[13] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and data fusion*. YBS publishing Storrs, CT, USA: 2011, vol. 11.

[14] H. Chen, T. Kirubarajan, and Y. Bar-Shalom, "Performance limits of track-to-track fusion versus centralized estimation: Theory and application [sensor fusion]," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 2, pp. 386–400, 2003.

[15] S. Sun, H. Lin, J. Ma, and X. Li, "Multi-sensor distributed fusion estimation with applications in networked systems: A review paper," *Information Fusion*, vol. 38, pp. 122–134, 2017.

[16] B. Bernhardsson and K. Åström, "Comparison of periodic and event based sampling for first-order stochastic systems," 14th IFAC World Congress (1999); Conference date: 05-07-1999 Through 09-07-1999, 1999.

[17] L. Grau, S. Dormido, C. Mañoso, and A. Pérez de Madrid, "F.M.S.C. An Adaptive Sampling System," *IFAC Proceedings Volumes*, vol. 28, no. 19, pp. 371–377, Sep. 1995.

[18] K. Åström and B. Bernhardsson, "Comparison of Riemann and Lebesgue sampling for first order stochastic systems," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 2, Las Vegas, NV, USA: IEEE, 2002, pp. 2011–2016.

[19] B. A. Moser and T. Natschlaeger, "On stability of distance measures for event sequences induced by level-crossing sampling," *IEEE Transactions on Signal Processing*, vol. 62, no. 8, pp. 1987–1999, 2014.

[20] P. Ellis, "Extension of phase plane analysis to quantized systems," *IRE Trans. Autom. Control*, vol. 4, no. 2, pp. 44–54, 1959.

[21] M. Lemmon, "Event-triggered feedback in control, estimation, and optimization," *Networked control systems*, pp. 293–358, 2010.

[22] Q. Liu, Z. Wang, X. He, and D. Zhou, "A survey of event-based strategies on control and estimation," *Systems Science & Control Engineering: An Open Access Journal*, vol. 2, no. 1, pp. 90–97, 2014.

[23] X. R. Li, "Optimal bayes joint decision and estimation," in *2007 10th International Conference on Information Fusion*, IEEE, 2007, pp. 1–8.

[24] M. Miskowicz, "Send-on-delta concept: An event-based data reporting strategy," *sensors*, vol. 6, no. 1, pp. 49–63, 2006.

[25] M Miskowicz, "The event-triggered integral criterion for sensor sampling," in *Proceedings of the IEEE International Symposium on Industrial Electronics, 2005. ISIE 2005.*, IEEE, vol. 3, 2005, pp. 1061–1066.

[26] P. G. Otanez, J. R. Moyne, and D. M. Tilbury, "Using deadbands to reduce communication in networked control systems," *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*, vol. 4, 3015–3020 vol.4, 2002.

[27] W. Heemels *et al.*, "Asynchronous measurement and control: A case study on motor synchronization," *Control Engineering Practice*, vol. 7, no. 12, pp. 1467–1482, 1999.

[28] Y. S. Suh, V. H. Nguyen, and Y. S. Ro, “Modified kalman filter for networked monitoring systems employing a send-on-delta method,” *Automatica*, vol. 43, no. 2, pp. 332–338, 2007.

[29] V. H. Nguyen and Y. S. Suh, “Improving estimation performance in networked control systems applying the send-on-delta transmission method,” *Sensors*, vol. 7, no. 10, pp. 2128–2138, 2007.

[30] Y. S. Suh, “Send-on-delta sensor data transmission with a linear predictor,” *Sensors (Basel, Switzerland)*, vol. 7, pp. 537 –547, 2007.

[31] V. H. Nguyen and Y. S. Suh, “Networked estimation with an area-triggered transmission method,” *Sensors*, vol. 8, no. 2, pp. 897–909, 2008.

[32] J. W. Marck and J. Sijs, “Relevant sampling applied to event-based state-estimation,” in *2010 Fourth International Conference on Sensor Technologies and Applications*, IEEE, 2010, pp. 618–624.

[33] J. Sijs, *State estimation in networked systems*. Eindhoven: Technical University Eindhoven, 2012.

[34] J. Sijs, B. Noack, and U. D. Hanebeck, “Event-based state estimation with negative information,” in *Proceedings of the 16th international conference on information fusion*, IEEE, 2013, pp. 2192–2199.

[35] S. Trimpe and R. D’Andrea, “Event-based state estimation with variance-based triggering,” *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3266–3281, 2014.

[36] S. Trimpe and M. C. Campi, “On the choice of the event trigger in event-based estimation,” in *2015 International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, IEEE, 2015, pp. 1–8.

[37] B. D. Anderson and J. B. Moore, *Optimal filtering*. Courier Corporation, 2012.

[38] E. J. Msechu, S. I. Roumeliotis, A. Ribeiro, and G. B. Giannakis, “Decentralized quantized kalman filtering with scalable communication cost,” *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3727–3741, 2008.

[39] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, “SOI-KF: Distributed kalman filtering with low-cost communications using the sign of innovations,” *IEEE Transactions on signal processing*, vol. 54, no. 12, pp. 4782–4795, 2006.

[40] J. Wu, Q.-S. Jia, K. H. Johansson, and L. Shi, “Event-based sensor data scheduling: Tradeoff between communication rate and estimation quality,” *IEEE Transactions on automatic control*, vol. 58, no. 4, pp. 1041–1046, 2012.

[41] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi, "Stochastic event-triggered sensor schedule for remote state estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 10, pp. 2661–2675, 2015.

[42] E. J. Schmitt, B. Noack, W. Krippner, and U. D. Hanebeck, "Gaussianity-preserving event-based state estimation with an FIR-based stochastic trigger," *IEEE Control Systems Letters*, vol. 3, no. 3, pp. 769–774, 2019.

[43] B. Noack, C. Öhl, and U. D. Hanebeck, "Event-based kalman filtering exploiting correlated trigger information," in *2022 25th International Conference on Information Fusion (FUSION)*, IEEE, 2022, pp. 1–8.

[44] M. T. Andrén and A. Cervin, "Event-based state estimation using an improved stochastic send-on-delta sampling scheme," in *2016 Second International Conference on Event-based Control, Communication, and Signal Processing (EBCCSP)*, IEEE, 2016, pp. 1–8.

[45] Y. Xu and J. P. Hespanha, "Estimation under uncontrolled and controlled communications in networked control systems," in *Proceedings of the 44th IEEE Conference on Decision and Control*, IEEE, 2005, pp. 842–847.

[46] J. Wu, Y. Yuan, H. Zhang, and L. Shi, "How can online schedules improve communication and estimation tradeoff?" *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1625–1631, 2013.

[47] J. Sijs, L. Kester, and B. Noack, "A study on event triggering criteria for estimation," in *17th International Conference on Information Fusion (FUSION)*, IEEE, 2014, pp. 1–8.

[48] D. Shi, T. Chen, and L. Shi, "Event-triggered maximum likelihood state estimation," *Automatica*, vol. 50, no. 1, pp. 247–254, 2014.

[49] K. You and L. Xie, "Kalman filtering with scheduled measurements," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1520–1530, 2013.

[50] Q. Liu, Z. Wang, X. He, and D.-H. Zhou, "Event-based recursive distributed filtering over wireless sensor networks," *IEEE Transactions on Automatic Control*, vol. 60, no. 9, pp. 2470–2475, 2015.

[51] X. Ge, Q.-L. Han, X.-M. Zhang, L. Ding, and F. Yang, "Distributed event-triggered estimation over sensor networks: A survey," *IEEE transactions on cybernetics*, vol. 50, no. 3, pp. 1306–1320, 2019.

[52] A. Girard, "Dynamic triggering mechanisms for event-triggered control," *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1992–1997, 2014.

[53] X. Ge, Q.-L. Han, X.-M. Zhang, and D. Ding, "Dynamic event-triggered control and estimation: A survey," *International Journal of Automation and Computing*, vol. 18, no. 6, pp. 857–886, 2021.

[54] W. Song, Z. Wang, J. Wang, F. E. Alsaadi, and J. Shan, "Distributed auxiliary particle filtering with diffusion strategy for target tracking: A dynamic event-triggered approach," *IEEE Transactions on Signal Processing*, vol. 69, pp. 328–340, 2020.

[55] Y. Liu, B. Shen, and H. Shu, "Finite-time resilient H_∞ state estimation for discrete-time delayed neural networks under dynamic event-triggered mechanism," *Neural Networks*, vol. 121, pp. 356–365, 2020.

[56] J. Cheng, L. Liang, J. H. Park, H. Yan, and K. Li, "A dynamic event-triggered approach to state estimation for switched memristive neural networks with nonhomogeneous sojourn probabilities," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 12, pp. 4924–4934, 2021.

[57] L. Wang, Z. Wang, G. Wei, and F. E. Alsaadi, "Finite-time state estimation for recurrent delayed neural networks with component-based event-triggering protocol," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 4, pp. 1046–1057, 2017.

[58] D. Shi, T. Chen, and L. Shi, "On set-valued kalman filtering and its application to event-based state estimation," *IEEE Transactions on Automatic Control*, vol. 60, no. 5, pp. 1275–1290, 2014.

[59] J. Sijs and M. Lazar, "Event based state estimation with time synchronous updates," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2650–2655, 2012.

[60] L. He, J. Chen, and Y. Qi, "Event-based state estimation: Optimal algorithm with generalized closed skew normal distribution," *IEEE Transactions on Automatic Control*, vol. 64, no. 1, pp. 321–328, 2018.

[61] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing for hidden markov models," *Systems & control letters*, vol. 25, no. 5, pp. 361–366, 1995.

[62] S. Dey and J. B. Moore, "Risk-sensitive filtering and smoothing via reference probability methods," *IEEE Transactions on Automatic Control*, vol. 42, no. 11, pp. 1587–1591, 1997.

[63] R. K. Boel, M. R. James, and I. R. Petersen, "Robustness and risk-sensitive filtering," *IEEE Transactions on Automatic Control*, vol. 47, no. 3, pp. 451–461, 2002.

[64] X.-M. Zhang and Q.-L. Han, "Event-based H_∞ filtering for sampled-data systems," *Automatica*, vol. 51, pp. 55–69, 2015.

[65] J. Xu, Y. Tang, W. Yang, F. Li, and L. Shi, “Event-triggered minimax state estimation with a relative entropy constraint,” *Automatica*, vol. 110, p. 108 592, 2019.

[66] J. Xu, D. W. Ho, F. Li, W. Yang, and Y. Tang, “Event-triggered risk-sensitive state estimation for hidden markov models,” *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 4276–4283, 2019.

[67] H. Liu, Z. Wang, B. Shen, and X. Liu, “Event-triggered H_∞ state estimation for delayed stochastic memristive neural networks with missing measurements: The discrete time case,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3726–3737, 2017.

[68] N. Li, Q. Li, and J. Suo, “Dynamic event-triggered H_∞ state estimation for delayed complex networks with randomly occurring nonlinearities,” *Neurocomputing*, vol. 421, pp. 97–104, 2021.

[69] S. Hu and D. Yue, “Event-based H_∞ filtering for networked system with communication delay,” *Signal Processing*, vol. 92, no. 9, pp. 2029–2039, 2012.

[70] X. Ge, Q.-L. Han, and Z. Wang, “A dynamic event-triggered transmission scheme for distributed set-membership estimation over wireless sensor networks,” *IEEE transactions on cybernetics*, vol. 49, no. 1, pp. 171–183, 2017.

[71] N. Xia, F. Yang, and Q.-L. Han, “Distributed event-triggered networked set-membership filtering with partial information transmission,” *IET Control Theory & Applications*, vol. 11, no. 2, pp. 155–163, 2017.

[72] L. Zou, Z. Wang, and D. Zhou, “Moving horizon estimation with non-uniform sampling under component-based dynamic event-triggered transmission,” *Automatica*, vol. 120, p. 109 154, 2020.

[73] J. Huang, D. Shi, and T. Chen, “Energy-based event-triggered state estimation for hidden markov models,” *Automatica*, vol. 79, pp. 256–264, 2017.

[74] A. S. Leong, S. Dey, and D. E. Quevedo, “Optimal transmission policies for variance based event triggered estimation with an energy harvesting sensor,” in *2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, 2016, pp. 225–229.

[75] J. Huang, D. Shi, and T. Chen, “Event-triggered state estimation with an energy harvesting sensor,” *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4768–4775, 2017.

[76] Y. Yılmaz, M. N. Kurt, and X. Wang, “Distributed dynamic state estimation and LQG control in resource-constrained networks,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, no. 3, pp. 599–612, 2018.

[77] G. Battistelli, A. Benavoli, and L. Chisci, "Data-driven communication for state estimation with sensor networks," *Automatica*, vol. 48, no. 5, pp. 926–935, 2012.

[78] G. Battistelli, L. Chisci, and D. Selvi, "A distributed kalman filter with event-triggered communication and guaranteed stability," *Automatica*, vol. 93, pp. 75–82, 2018.

[79] S. Trimpe, "Stability analysis of distributed event-based state estimation," in *53rd IEEE Conference on Decision and Control*, IEEE, 2014, pp. 2013–2019.

[80] S. Weerakkody, Y. Mo, B. Sinopoli, D. Han, and L. Shi, "Multi-sensor scheduling for state estimation with event-based, stochastic triggers," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2695–2701, 2015.

[81] D. Shi, T. Chen, and L. Shi, "An event-triggered approach to state estimation with multiple point-and set-valued measurements," *Automatica*, vol. 50, no. 6, pp. 1641–1648, 2014.

[82] A. Molin, H. Sandberg, and K. H. Johansson, "Consistency-preserving event-triggered estimation in sensor networks," in *2015 54th IEEE Conference on Decision and Control (CDC)*, IEEE, 2015, pp. 7494–7501.

[83] A. S. Leong, S. Dey, and D. E. Quevedo, "Sensor scheduling in variance based event triggered estimation with packet drops," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 1880–1895, 2016.

[84] A. Mohammadi and K. N. Plataniotis, "Event-based estimation with information-based triggering and adaptive update," *IEEE Transactions on Signal Processing*, vol. 65, no. 18, pp. 4924–4939, 2017.

[85] Z. Hu, B. Chen, W.-A. Zhang, and L. Yu, "Event-triggered fusion estimation under bounded noises," *Automatica*, vol. 154, p. 111 060, 2023.

[86] K. Shen, C. Zhang, P. Dong, Z. Jing, and H. Leung, "Consensus-based labeled multi-bernoulli filter with event-triggered communication," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1185–1196, 2022.

[87] M. Muehlebach and S. Trimpe, "Distributed event-based state estimation for networked systems: An LMI approach," *IEEE Transactions on Automatic Control*, vol. 63, no. 1, pp. 269–276, 2017.

[88] W. Li, Y. Jia, and J. Du, "Event-triggered kalman consensus filter over sensor networks," *IET Control Theory & Applications*, vol. 10, no. 1, pp. 103–110, 2016.

[89] Y. Ju, D. Ding, X. He, Q.-L. Han, and G. Wei, "Consensus control of multi-agent systems using fault-estimation-in-the-loop: Dynamic event-triggered case," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 8, pp. 1440–1451, 2021.

[90] L. Ma, Z. Wang, H.-K. Lam, and N. Kyriakoulis, “Distributed event-based set-membership filtering for a class of nonlinear systems with sensor saturations over sensor networks,” *IEEE transactions on cybernetics*, vol. 47, no. 11, pp. 3772–3783, 2016.

[91] J. Weimer, J. Araújo, and K. H. Johansson, “Distributed event-triggered estimation in networked systems,” *IFAC Proceedings Volumes*, vol. 45, no. 9, pp. 178–185, 2012.

[92] O. C. Imer and T. Basar, “Optimal estimation with limited measurements,” in *Proceedings of the 44th IEEE Conference on Decision and Control*, IEEE, 2005, pp. 1029–1034.

[93] M. Rabi, G. V. Moustakides, and J. S. Baras, “Multiple sampling for estimation on a finite horizon,” in *Proceedings of the 45th IEEE Conference on Decision and Control*, IEEE, 2006, pp. 1351–1357.

[94] L. Li, M. Lemmon, and X. Wang, “Event-triggered state estimation in vector linear processes,” in *Proceedings of the 2010 American control conference*, IEEE, 2010, pp. 2138–2143.

[95] J. Wu, K. H. Johansson, and L. Shi, “A stochastic online sensor scheduler for remote state estimation with time-out condition,” *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 3110–3116, 2014.

[96] Y. Xu and J. P. Hespanha, “Optimal communication logics in networked control systems,” in *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, IEEE, vol. 4, 2004, pp. 3527–3532.

[97] A. Molin and S. Hirche, “Event-triggered state estimation: An iterative algorithm and optimality properties,” *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5939–5946, 2017.

[98] A. S. Leong, S. Dey, and D. E. Quevedo, “On the optimality of threshold policies in event triggered estimation with packet drops,” in *2015 European Control Conference (ECC)*, IEEE, 2015, pp. 927–933.

[99] J. Wu, X. Ren, D. Han, D. Shi, and L. Shi, “Finite-horizon gaussianity-preserving event-based sensor scheduling in kalman filter applications,” *Automatica*, vol. 72, pp. 100–107, 2016.

[100] L. Xu, Y. Mo, and L. Xie, “Remote state estimation with stochastic event-triggered sensor schedule and packet drops,” *IEEE Transactions on Automatic Control*, vol. 65, no. 11, pp. 4981–4988, 2020.

[101] Z. Jin, Y. Hu, and C. Sun, “Event-triggered information fusion for networked systems with missing measurements and correlated noises,” *Neurocomputing*, vol. 332, pp. 15–28, 2019.

[102] L. Wang, Z. Wang, T. Huang, and G. Wei, "An event-triggered approach to state estimation for a class of complex networks with mixed time delays and nonlinearities," *IEEE Transactions on Cybernetics*, vol. 46, no. 11, pp. 2497–2508, 2015.

[103] L. Ma, Z. Wang, C. Cai, and F. E. Alsaadi, "Dynamic event-triggered state estimation for discrete-time singularly perturbed systems with distributed time-delays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 9, pp. 3258–3268, 2018.

[104] Q. Liu, Z. Wang, X. He, and D. Zhou, "Event-triggered resilient filtering with measurement quantization and random sensor failures: Monotonicity and convergence," *Automatica*, vol. 94, pp. 458–464, 2018.

[105] D. Deng and J. Xiong, "Stochastic event-triggered remote state estimation over gaussian channels without knowing triggering decisions: A bayesian inference approach," *Automatica*, vol. 152, p. 110 951, 2023.

[106] J. Liu, L. Wei, X. Xie, and D. Yue, "Distributed event-triggered state estimators design for sensor networked systems with deception attacks," *IET Control Theory & Applications*, vol. 13, no. 17, pp. 2783–2791, 2019.

[107] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE transactions on automatic control*, vol. 45, no. 5, pp. 910–927, 2000.

[108] X. Bian and Y. Xia, "Energy-efficient data fusion over wireless channels with power control," *IET Signal Processing*, vol. 9, no. 3, pp. 206–217, 2015.

[109] G. Tallis, "Elliptical and radial truncation in normal populations," *The Annals of Mathematical Statistics*, vol. 34, no. 3, pp. 940–944, 1963.

[110] X. R. Li, "Tracking in clutter with strongest neighbor measurements. i. theoretical analysis," *IEEE Transactions on Automatic Control*, vol. 43, no. 11, pp. 1560–1578, 1998.

[111] E. R. Ziegel, "Standard probability and statistics tables and formulae," *Technometrics*, vol. 43, no. 2, p. 249, 2001.

[112] L. Shi, M. Epstein, and R. M. Murray, "Kalman filtering over a packet-dropping network: A probabilistic perspective," *IEEE Transactions on Automatic Control*, vol. 55, no. 3, pp. 594–604, 2010.

[113] D. S. Bernstein, "Matrix mathematics," in *Matrix Mathematics*, Princeton university press, 2009.

[114] E. Grigorieva, *Methods of solving sequence and series problems*. Springer, 2016.

[115] X. R. Li and V. P. Jilkov, “Survey of maneuvering target tracking. part i. dynamic models,” *IEEE Transactions on aerospace and electronic systems*, vol. 39, no. 4, pp. 1333–1364, 2003.

[116] D. Bertsekas, *Dynamic programming and optimal control: Volume II*. Athena scientific, 1995, vol. 2.

[117] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 1994.

[118] D. Bertsekas, *Dynamic programming and optimal control: Volume II*. Athena scientific, 2012, vol. 2.