2016

JOINT DETECTION-STATE ESTIMATION AND SECURE SIGNAL PROCESSING

Mengqi Ren

Virginia Commonwealth University

JOINT DETECTION-STATE ESTIMATION AND SECURE SIGNAL PROCESSING

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

MENGQI REN

M.S., Peking University, China - September 2009 to July 2012

B.E. & B.S., East China Normal University, China - September 2005 to July 2009

Adviser: Ruixin Niu, Ph.D.,

Assistant Professor, Department of Electrical and Computer Engineering

Virginia Commonwealth University

Richmond, Virginia

December, 2016

ACKNOWLEDGMENTS

The author wishes to thank several people. First and foremost I offer my sincerest gratitude to my advisor, Dr. Ruixin Niu, who has supported me throughout my dissertation with his patience, knowledge, and motivation. I attribute the level of my doctorate degree to his encouragement and rigorous approach to scholarly research. Without his guidance and constant feedback this dissertation would not have been finished.

In addition, I would like to thank the other members of my doctoral advisory committee: Dr. Alen Docef, Dr. Yuichi Motai, Dr. Qiqi Lu, and Dr. Hong-sheng Zhou, for their encouragement, assistance, insightful comments, and suggestions.

I would also like to thank all of my family for their continued love and support through my entire life. A special thank you goes to my paternal grandmother and maternal grandmother for their love and sacrifice while I study in the U.S. This dissertation is for them.

Finally, I would like to thank all my friends for their help, friendship, and support in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

JOINT DETECTION-STATE ESTIMATION AND SECURE SIGNAL PROCESSING

By Mengqi Ren, Ph.D.

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2016.

Advisor: Ruixin Niu, Ph.D.,

Assistant Professor, Department of Electrical and Computer Engineering

In this dissertation, joint detection-state estimation and secure signal processing are studied. Detection and state estimation are two important research topics in surveillance systems. The detection problems investigated in this dissertation include object detection and fault detection. The goal of object detection is to determine the presence or absence of an object under measurement uncertainty. The aim of fault detection is to determine whether or not the measurements are provided by faulty sensors. State estimation is to estimate the states of moving objects from measurements with random measurement noise or disturbance, which typically consist of their positions and velocities over time. Detection and state estimation are typically implemented separately and state estimation is usually performed after the decision is made. In this two-stage approach, missed detection and false alarms in detection stage decrease accuracy of state estimation. In this dissertation, several joint detection and state estimation algorithms are proposed. Secure signal processing is indispensable in dynamic systems especially when an adversary exists. In this dissertation, the

developed joint fault detection and state estimation approach is used to detect attacks launched by an adversary on the system and improve state estimation accuracy. The security problem in satellite communication systems is studied and a minimax anti-jammer is designed in a frequency hopping spread spectrum (FHSS)/quadrature phase-shift keying (QPSK) satellite communication system.

In object detection problems, the false alarms create ghost targets, especially when the object has a very low signal to noise ratio (SNR) and a decision is made with a single sample. If the system state is estimated after an object is detected, the presence of ghost target(s) decreases the accuracy of the system state estimate. In this dissertation, a new joint sequential object detection and system state estimation algorithm is proposed, which has the potential to significantly improve the detection of extremely weak targets or phenomena. This algorithm combines Wald's sequential probability ratio test (SPRT) and the Kalman filter. Theoretical results have been provided on the first and second moments of the test statistic under both hypotheses, to calculate Kullback-Leibler distance and deflection coefficient, and give insights on the termination of the SPRT procedure. To guarantee the termination of the sequential test with probability one, a joint terminative sequential detection and system state estimation algorithm is proposed, which uses fused test statistic in Wald's SPRT. A method of choosing the thresholds by using the nominal probabilities of error is proposed. The upper bounds on the actual probabilities of error are derived.

The problem of faulty sensor detection is investigated in large sensor networks where sensors' measurement noise is correlated and the faulty sensors are sparse and time-varying. Similar to object detection and state estimation, fault detection and state estimation are usually implemented separately. However, when faulty sensors are missed in fault detection, their measurements will deteriorate the state estimation performance. Furthermore, when sensors' measurement noise is correlated, detecting

faulty sensors by testing each single sensor individually is not optimal. To improve the performance of detection and state estimation, an approach for joint group testing of time-varying faulty sensors and state estimation in the presence of correlated measurement noise is proposed.

In satellite communication systems, jamming attack from an adversary is a significant threat. A jammer can simply interfere the legitimate communication system by injecting jamming signals into communication channels. As a result, the communication between the transmitter and receiver will be corrupted. To achieve reliable communication, employing anti-jammer is crucial in satellite communication systems. One efficient way to design anti-jammer is adopting FHSS, which avoids attacks by switching channels from time to time. In this dissertation, a minimax anti-jamming strategy in a FHSS/QPSK satellite communication system is developed.

CHAPTER 1

INTRODUCTION

In this dissertation, joint detection-state estimation and secure signal processing are studied, in which both object detection and faulty sensor detection are investigated. In surveillance systems, object detection and system state estimation are two important problems. They are typically implemented separately and system state estimation is implemented after an object is detected. However, this two-stage approach may generate ghost targets because of false alarms and/or fail to detect the target, especially when the object has a very low signal to noise ratio (SNR) and a decision is made with a single sample. The false alarms and/or missed detection in detection stage will also decrease the accuracy of system state estimation. That motivates the research on joint detection-estimation to improve both the detection and estimation performance by integrating information over multiple samples using a system state estimator.

In large sensor networks, faulty sensor detection is crucial in system reliability, especially when the sensors are under the risk of attacks. If the measurements obtained by all the sensors are used to estimate the system state, the accuracy of system state estimation will be decreased by the existence of faulty sensors. Detecting faulty sensors before estimating the system state is one possible solution, but the miss in the detection stage will still affect the performance of system state estimation. Furthermore, testing sensors one by one is not optimal when measurement noise is correlated. So, joint detection-estimation is also studied in faulty sensor detection and system state estimation.

Besides in surveillance systems, the security issue is also important in satellite communication systems. A jammer can simply interfere the legitimate communication systems by injecting jamming signals into communication channels [1, 2]. As a result, the communication between the transmitter and receiver will be corrupted. To achieve reliable communication, employing anti-jammer is crucial in satellite communication systems.

In this dissertation, four algorithms have been proposed: a new joint sequential object detection and system state estimation approach based on Wald's sequential probability ratio test (SPRT), terminative joint sequential object detection and system state estimation based on fused test statistics, joint group testing of time-varying faulty sensors and system state estimation with correlated measurement noise, and minimax anti-jammer design for frequency hopping spread spectrum (FHSS)/ quadrature phase-shift keying (QPSK) satellite communication systems. Detection and system state estimation are two important problems in the proposed algorithms. The detection mentioned here includes object detection and faulty sensor detection. Object detection is to determine the presence or absence of an object under uncertainty. The aim of fault detection is to determine whether the measurements are corrupted or not under uncertainty, where the corruption could be caused by injected false information by an adversary, abnormal behavior of sensor, etc. [3]. In detection algorithms, binary hypothesis testing is widely applied, in which the general or default status is usually represented by null hypothesis and the status needs to be detected is usually represented by alternative hypothesis. In object detection, the absence and presence of an object are usually represented by null hypothesis and alternative hypothesis, respectively. In fault sensor detection, the absence and presence of faulty sensor(s) are usually represented by null hypothesis and alternative hypothesis, respectively. The goal of system state estimation is to estimate the states of moving objects from

inaccurate and uncertain measurements, which typically consist of their positions and velocities over time. The estimation of states of dynamic system is termed as filtering, because the best estimate is obtained from noisy measurements via filtering out the noise [4]. The optimal estimator for linear dynamic system with additive Gaussian noise is Kalman filter [5] which is a minimum mean squared error estimator. If the system is linear and Gaussian, only conditional mean and covariance need to be estimated recursively. However, in the nonlinear case, the conditional mean is obtained by updating the posterior probability density function (PDF) of the system state recursively, in which the accuracy of the estimator is measurement-dependent and the computational complexity is exponential in the dimension of the state [4]. Therefore, some suboptimal recursive Bayesian filters are widely applied, which include extended Kalman Filter [4], unscented Kalman Filter [6], particle filter [7], and particle flow filter [8]. In the following sections, some classical algorithms of detection and system state estimation are presented.

1.1 Detection Algorithms

Detection is to decide when an event of interest occurs. The technology to achieve this goal is called detection theory in the context of radar and communication, and it is well known as hypothesis testing or decision theory in statistics [9]. The simplest detection problem is to decide whether a signal is present combined with noise or only noise is present. This detection problem can be modeled as binary hypothesis testing problem.

Hypothesis testing is a standard procedure for testing a claim or statement about a property of a population. A hypothesis test includes four components: null hypothesis (denoted by H_0), alternative hypothesis (denoted by H_1), test statistic, and critical region (or rejection region). Null hypothesis is a statement that the value of a popula-

tion parameter is equal to some claimed value. The null hypothesis is assumed to be true and to be tested. In the simplest detection problem, null hypothesis is noise only hypothesis. Alternative hypothesis is a statement that the parameter has a value that differs somewhat from the null hypothesis. In the simplest detection problem, there is only one alternative hypothesis and it is signal with noise hypothesis. Test statistic is a sample statistic (a value) used in making a decision about the null hypothesis. Critical region is the range of values of the test statistic that will lead to reject the null hypothesis. The hypothesis test leads to the conclusion that either rejects or fails to reject null hypothesis. Therefore, the potential outcomes can be classified into four categories: rejecting the null hypothesis when it is actually false (which is called hit or detection in detection theory), rejecting the null hypothesis when it is actually true (which is called false alarm in detection theory and type I error in hypothesis testing), failing to reject the null hypothesis when it is actually false (which is called miss in detection theory and type II error in hypothesis testing), failing to reject the null hypothesis when it is actually true (which is called correct rejection in detection theory). A common way to display these outcomes is shown in Table 1.

Table 1. Four categories of potential outcomes of hypothesis testing

		True State of Signal	
		Present (H_1)	Absent (H_0)
Decision	Present (H_1)	Detection/Hit	False Alarm
	Absent (H_0)	Miss	Correct Rejection

The power of a test is defined as the probability of rejecting H_0 [10] which can

be expressed as

$$Power(\theta) = P_\theta(\text{rejecting } H_0) = \begin{cases} \alpha & \text{if } H_0 \text{ is true} \\ 1-\beta & \text{if } H_1 \text{ is true} \end{cases} \tag{1.1}$$

where θ denotes parameter(s) of distribution of the random variable(s), α and β denote probability of false alarm and probability of miss respectively.

Generally, the tests which minimize probabilities of error are optimal. By controlling probability of false alarm, most powerful test is the optimal test. Neyman-Pearson test is a most powerful (or an optimal) test [9] since it minimizes the probability of miss by controlling the probability of false alarm. Neyman-Pearson test (also known as likelihood ratio test) is shown in Theorem 1.

Theorem 1 *Denote a set of measurements as* $\mathbf{z}_{1:K} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_K\}$*, where* K *is a fixed number. Let the likelihood of* $\mathbf{z}_{1:K}$ *under* H_0 *and* H_1 *be* $p(\mathbf{z}_{1:K}|H_0)$ *and* $p(\mathbf{z}_{1:K}|H_1)$ *respectively. To maximize probability of detection* P_D *for a given probability of false alarm* $P_{FA} = \alpha$*, decide* H_1 *if the likelihood ratio satisfies the following inequality; otherwise, decide* H_0

$$L(\mathbf{z}_{1:K}) = \frac{p(\mathbf{z}_{1:K}|H_1)}{p(\mathbf{z}_{1:K}|H_0)} > \gamma \tag{1.2}$$

where the threshold γ *is found from*

$$P_{FA} = \int_{\{\mathbf{z}_{1:K}: L(\mathbf{z}_{1:K}) > \gamma\}} p(\mathbf{z}_{1:K}|H_0) d\mathbf{z}_{1:K} = \alpha$$

Neyman-Pearson is not the only optimal test. To specify clearly, the test procedures are classified into two categories according to whether the sample size is fixed or not. If the decision is made based on pre-defined number of samples, the procedure is called fixed-sample-size (FSS) procedure. Neyman-Pearson test is one FSS procedure. Contrary to FSS procedures, the number of samples needed by sequential hypothesis test procedures is a random variable. The sequential hypothesis tests

include sequential probability ratio test (SPRT), generalized sequential probability ratio test (GSPRT) [11], and Chernoff test [12]. The advantage of SPRT is that its required average (expected) sample number (ASN) is smaller than that required by the FSS procedure to achieve the same detection performance.

Wald's SPRT is a sequential procedure. At each stage of the experiment, the likelihood ratio for this stage is compared with two thresholds. If it is between the two thresholds, the test enters next stage and the likelihood ratio will be recalculated by combining next sample. Wald's SPRT procedure will continue this iteration until the likelihood ratio passes either threshold. The Wald's SPRT is summarized as follows [13]:

$$\frac{p_K(\mathbf{z}_{1:K}|H_1)}{p_K(\mathbf{z}_{1:K}|H_0)} \begin{cases} \geq A & \text{stop and decide } H_1 \\ \leq B & \text{stop and decide } H_0 \\ \text{otherwise} & \text{continue} \end{cases} \tag{1.3}$$

where K denotes the number of samples which increases stage by stage until either H_1 or H_0 is decided, A and B are two positive constants and $B < A$, which are determined by pre-specified probabilities of false alarm and miss.

In Wald's SPRT, the thresholds A and B, actual probability of false alarm α^a, and actual probability of miss β^a satisfy the following inequalities [13].

$$A \leqslant \frac{1-\beta^a}{\alpha^a} \tag{1.4}$$

$$B \geqslant \frac{\beta^a}{1-\alpha^a} \tag{1.5}$$

Therefore, the actual probabilities of false alarm and miss, i.e., α^a and β^a, are controlled by the thresholds A and B. Let α^n and β^n be the nominal probabilities of false alarm and miss, respectively. Since $0 \leqslant \alpha^a \leqslant 1$ and $0 \leqslant \beta^a \leqslant 1$, if $A = \frac{1-\beta^n}{\alpha^n}$ and $B = \frac{\beta^n}{1-\alpha^n}$, then actual probabilities of error and nominal probabilities of error

satisfy the following inequalities.

$$\alpha^a \leqslant \frac{\alpha^n}{1-\beta^n} \tag{1.6}$$

and

$$\beta^a \leqslant \frac{\beta^n}{1-\alpha^n} \tag{1.7}$$

Wald's SPRT satisfies Wald-Wolfowitz optimality property when the samples are independent and identically distributed (i.i.d.) random variables [14]. The Wald-Wolfowitz optimality property is defined as neither probability of false alarm nor probability of miss of an SPRT can be reduced without recourse to a test of larger expected sample size under both hypotheses.

Besides Wald's SPRT, there exist other SPRTs which also satisfy the Wald-Wolfowitz optimality property under the i.i.d. model [15]. The SPRT is summarized as follows:

$$\frac{p_K(\mathbf{z}_{1:K}|H_1)}{p_K(\mathbf{z}_{1:K}|H_0)} \begin{cases} \geq A_K & \text{stop and decide } H_1 \\ \leq B_K & \text{stop and decide } H_0 \\ \text{otherwise} & \text{continue} \end{cases} \tag{1.8}$$

where A_K and B_K are the thresholds that are functions of K and $0 \leqslant B_k < A_K \leqslant \infty$ for $K = 1, 2, \ldots$.

If $0 \leqslant B_k \leqslant A_K \leqslant \infty$ for all K, the SPRT test becomes GSPRT which is optimal when the samples are independent but not identically distributed [16]. Comparing with SPRT, there is one additional case in GSPRT which is $A_K = B_K$. When $\frac{p_K(\mathbf{z}_{1:K}|H_1)}{p_K(\mathbf{z}_{1:K}|H_0)} = A_K = B_K$, the decision is made according to the following rules: H_0 is decided if $\frac{p_K(\mathbf{z}_{1:K}|H_1)}{p_K(\mathbf{z}_{1:K}|H_0)} = A_K = B_K = 0$, H_1 is decided if $\frac{p_K(\mathbf{z}_{1:K}|H_1)}{p_K(\mathbf{z}_{1:K}|H_0)} = A_K = B_K = \infty$, and if $0 < \frac{p_K(\mathbf{z}_{1:K}|H_1)}{p_K(\mathbf{z}_{1:K}|H_0)} = A_K = B_K < \infty$, either H_0 or H_1 can be decided by randomization.

When the samples are dependent, the SPRT and GSPRT is optimal by intro-

ducing weakly admissibility which is a weaker notion of optimality [15]. A sequential test for deciding between H_0 and H_1, whose stopping variable is N_s and whose error probabilities are α and β, will be called inadmissible if there exists an alternative test T', with stopping variable N'_s and error probabilities α' and β', such that

$$\alpha' \leqslant \alpha, \quad P(N'_s > K|H_0) \leqslant P(N_s > K|H_0),$$
$$\beta' \leqslant \beta, \quad P(N'_s > K|H_1) \leqslant P(N_s > K|H_1), \quad K = 1, 2, \ldots$$

with at least one of these inequalities strict. T is admissible if no such T' exists [15, 17]. A test T will be called weakly admissible if there is no alternative test T' which not only improves upon T in the sense of admissibility but also stops no later than T does under H_0 and H_1. It was shown that every SPRT is weakly admissible, and a GSPRT is weakly admissible if $0 \leqslant B_{K+1} \leqslant B_K \leqslant A_K \leqslant A_{K+1} \leqslant \infty$ for each K.

Another optimal detector is Bayesian detector, which is suitable for the case where the hypotheses H_0 and H_1 have certain prior probabilities. The objective of the detector is to minimize the expected cost which is called Bayes risk. When the costs assigned to detection and correct rejection are zeros and the costs assigned to false alarm and miss are ones, the objective of the detector becomes to minimize the probability of error. The test statistic used in a Bayesian detector is also the likelihood ratio. The threshold is determined by the prior probabilities of the two hypotheses. In this dissertation, the focus will be on non-Bayesian sequential detection procedures.

1.2 System State Estimation Algorithms

The Kalman filter is the optimal estimator for linear dynamic system with additive Gaussian noise. Consider a discrete-time linear dynamic system given by the following state-space model [4]

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{\Gamma}\mathbf{v}_k \tag{1.9}$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k \tag{1.10}$$

where $\mathbf{x}_k$ is $n_x \times 1$ state vector at time k, $\mathbf{z}_k$ is the $n_z \times 1$ measurement vector at time k, $\mathbf{F}$ is $n_x \times n_x$ state transition matrix, $\mathbf{H}$ is the $n_z \times n_x$ measurement matrix, $\mathbf{w}_k$ is the measurement noise at time k, $\mathbf{v}_k$ is the process noise at time k, and $\mathbf{\Gamma}$ is the gain matrix for $\mathbf{v}_k$. The $\{\mathbf{v}_k\}$ and $\{\mathbf{w}_k\}$ are sequences of white Gaussian process noise with zero-mean and covariance matrices $\mathbf{Q}$ and $\mathbf{R}_w$, respectively. Furthermore, $\{\mathbf{x}_0\}$, $\{\mathbf{v}_k\}$ and $\{\mathbf{w}_k\}$ are independent.

The Kalman filter consists of two steps at each time: prediction and update. Since the state-space model is linear and Gaussian, all the random variables follow Gaussian distributions which can be determined by their means and variances. In the prediction step, the conditional distribution of $\mathbf{x}_{k+1}$ is predicted. The mean of $\mathbf{x}_{k+1}$ conditioned on $\mathbf{z}_{1:k}$ is termed as state prediction, which is obtained as follows

$$\hat{\mathbf{x}}_{k+1|k} = E(\mathbf{x}_{k+1}|\mathbf{z}_{1:k}) = \mathbf{F}\hat{\mathbf{x}}_{k|k} \tag{1.11}$$

The state prediction covariance which is covariance matrix of $\mathbf{x}_{k+1}$ conditioned on $\mathbf{z}_{1:k}$ is given as

$$\mathbf{P}_{k+1|k} = \mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^T + \mathbf{\Gamma}\mathbf{Q}\mathbf{\Gamma}^T \tag{1.12}$$

In the update step, measurement prediction of $\mathbf{z}_{k+1}$ is calculated first.

$$\hat{\mathbf{z}}_{k+1|k} = \mathbf{H}\hat{\mathbf{x}}_{k+1|k} \tag{1.13}$$

The measurement prediction covariance is

$$\mathbf{S}_{k+1} = \mathbf{H}\mathbf{P}_{k+1|k}\mathbf{H}^T + \mathbf{R}_w \tag{1.14}$$

The updated state estimate is

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{W}_{k+1}\boldsymbol{\nu}_{k+1} \tag{1.15}$$

where $\boldsymbol{\nu}_{k+1} \triangleq \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|k}$ is called innovation or measurement residual, and $\mathbf{W}_{k+1}$ is the filter gain which is defined as follows

$$\mathbf{W}_{k+1} \triangleq \mathbf{P}_{k+1|k}\mathbf{H}^T\mathbf{S}_{k+1}^{-1} \tag{1.16}$$

The updated state covariance is

$$\mathbf{P}_{k+1|k+1} = \mathbf{P}_{k+1|k} - \mathbf{W}_{k+1}\mathbf{S}_{k+1}\mathbf{W}_{k+1}^T \tag{1.17}$$

The input of the Kalman filter is nonwhite measurement sequence $\{\mathbf{z}_k\}$, but output of the Kalman filter is the innovation which is a zero mean and white sequence. Therefore, the Kalman filter is a whitening system for the measurement sequence. Under the Gaussian assumption, the innovation is not only a zero mean and white sequence, but also follows Gaussian distribution and independent with each other. Furthermore, the PDF of innovation $\boldsymbol{\nu}_{k+1}$ can be obtained from the PDF of measurement $\mathbf{z}_{k+1}$ conditioned on $\mathbf{z}_{1:k}$ through a simple mean shift. These properties can be utilized to design joint detection and system state estimation approaches.

This dissertation is organized as follows. In Chapter 2, the problem of object detection and system state estimation is investigated, and a new joint sequential object detection and system state estimation approach based on Wald's SPRT is developed. In Chapter 3, terminative joint sequential object detection and system state estimation based on fused test statistics is developed to guarantee that the sequential test will eventually terminate with probability one. In Chapter 4, the problem of faulty sensor detection and system state estimation is investigated, and joint group testing of time-varying faulty sensors and system state estimation with correlated measurement noise is developed. In Chapter 5, the security problem in satellite communication systems is investigated, and a minimax anti-jammer is designed for FHSS/QPSK satellite communication systems. Finally, this dissertation is concluded in Chapter 6.

CHAPTER 2

A NEW JOINT SEQUENTIAL OBJECT DETECTION AND SYSTEM STATE ESTIMATION APPROACH

2.1 Introduction

For most surveillance systems, object detection and system state estimation are two important problems that need to be solved. The goal of object detection is to determine the presence or absence of a target under uncertainty. In object detection algorithms, the presence and absence of targets are usually represented by two hypotheses H_1 and H_0 respectively, under which the knowledge of distributions of measurements are required. System state estimation is to estimate the states of moving targets, which typically consist of their positions and velocities over time. The system state estimation algorithms usually assume the presence of the target(s). The object detection and system state estimation are typically implemented separately and system state estimation is performed after the target is detected, which is shown by the first diagram in Fig. 1. This two-stage approach works well when the target has a relatively high signal-to-noise ratio (SNR), and it can be reliably detected. But this approach may not detect the weak target reliably with acceptable detection performance using a single sample. This motivates the research on joint object detection and system state estimation, which has the potential to significantly improve the detection of extremely weak targets or phenomena, such as a weak target that is far away from the radar or the chemical/biological plumes with very low concentration. The joint object detection and system state estimation is compared with the first procedure in Fig. 1.

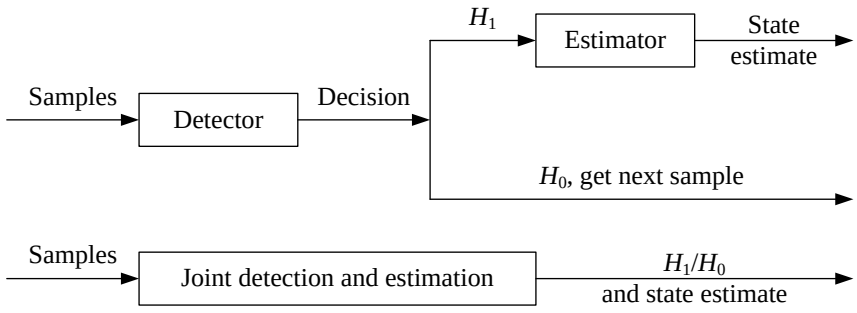

Fig. 1. Comparison of Two Detection and System State Estimation Procedures

There are a number of publications on joint detection and parameter estimation in the literature [18, 19, 20, 21, 22]. Using a Bayes criterion, the optimum detector and estimator structures were derived for joint detection and estimation by minimizing the average risk in [18]. Later, the results in [18], which were derived based on the binary-hypothesis assumption, were extended to the multi-hypothesis case in [19]. The trade-off between detection and estimation for a finite discrete parameter space was studied in [20]. The assumption of finite discrete parameter space makes it possible to convert the parameter estimation problem to a detection sub-problem. In [21], the joint detection and estimation problem was defined as an optimization problem in which the objective function is associated with the Bayesian formulation of the estimation sub-problem and the constraints depend on the detection sub-problem. The algorithm proposed in [21] assumes that only the parameters in the alternative hypothesis are unknown. This assumption was later relaxed in [22]. However, these algorithms cannot be directly applied in the joint detection and system state estimation problems, since in a system state estimation problem, the estimator needs to sequentially estimate the system state, which is time-varying.

Joint detection and system state estimation methods can be used to solve joint object detection and tracking problems. There are only a very limited number of pub-

lications on joint detection and tracking. One joint detection and tracking algorithm is called multiple multistage hypothesis test tracking algorithm [23]. This algorithm detects targets by evaluating candidate track hypotheses via test statistic in the multistage hypothesis testing algorithm [24]. The premise of this algorithm is that the possible target trajectories are known. One track-before-detect algorithm was proposed in [25], which utilizes a dynamic programming algorithm to evaluate candidate target states. This algorithm detects target by comparing the accumulated maximum target amplitude over time with one empirical threshold. The target trajectory is formed by the candidate state with maximum target amplitude at each time. This algorithm requires large data storage and needs to evaluate all the candidate states. In [26], truncated sequential probability ratio test (SPRT) was introduced into joint detection and tracking in which both likelihood ratio test (LRT) and maximum a posteriori (MAP) can be applied to detect target. Note that all the algorithms mentioned above assume the state space is discrete. In [27], multiple active radars were adopted to design joint detection and tracking algorithm. The decision was made by comparing ratio of probabilities of amplitude under two hypotheses with a Bayesian threshold which is obtained via the probabilities of predicted measurements in target tracking. Since tracking is activated when the decision is positive, the Bayesian threshold will not be updated if no target is detected in last step. In [28], a tracking algorithm without detection was proposed. A lidar sensor was used to measure the surrounding environment and returns environment information in polar coordinates. The system state estimate was obtained via a particle filter. The measurements used by the particle filter are the values of a cumulative function which measures the densities of the objects and the distances between the objects and the lidar sensor in each angular interval. In [29], an algorithm using the Bernoulli filter was proposed to perform joint detection and tracking. This algorithm assumes that the presence

or absence of a single target is time-varying and models it as a random set: the posterior probability density function (PDF) of the set's cardinality determines the presence/absence state of the target, and the posterior PDF of the element in the set corresponds to the target's state given the presence of target. The two posterior PDFs of the random set are updated recursively in a Bayesian framework. In this algorithm, the detection decision is made by comparing the posterior probabilities of the two hypotheses, and the hypothesis with higher posterior probability is decided.

Different from all the joint detection and system state estimation algorithms discussed above, a joint object detection and system state estimation algorithm based on the LRT and the extended Kalman filter (EKF) was proposed for sensor networks [30]. This algorithm has the following advantages: it works in continuous state space, applies the most powerful test — LRT, does not need the knowledge of possible system state trajectories, does not require very informative prior knowledge, does not require large data storage, and has acceptable computational complexity. However, this algorithm is a fixed-sample-size (FSS) procedure, where the number of samples has been pre-specified. It is well known that sequential detection on the average requires a smaller number of measurements than a detection procedure with a fixed sample size [13]. Moreover, SPRT is always weakly admissible [15], which will be discussed in detail later in the chapter. Therefore, to reduce the average sample number (ASN), we apply Wald's SPRT instead of LRT to perform joint detection and system state estimation. This feature is very attractive in radar surveillance systems, which try to declare the presence of a target as early as possible with acceptable detection performance.

In order to obtain useful insights and theoretical results, we study a general linear Gaussian system state estimation problem in this chapter. It was proved in [13] that Wald's SPRT procedure will eventually terminate with probability one if the

measurements are independent. However, the successive measurements are dependent in the proposed joint sequential object detection and system state estimation algorithm under hypothesis H_1, which means the existence of a target. In such a case, it is difficult to rigorously prove that the Wald's SPRT will eventually terminate with probability one. Instead, we analytically derive the expected value and variance of the test statistic under both hypotheses. For the example given in this chapter, we show that the expected value of the test statistic under the two hypotheses, which is proportional to the Kullback-Leibler distance, is either a monotonically increasing or decreasing function of the number of samples. This implies that most probably the Wald's SPRT will eventually terminate. The deflection coefficient is also applied to evaluate the termination and detection performance of the proposed algorithm. We show that the deflection coefficient is increasing over time, which means that the probability of termination is increasing with time and the detection performance is also increasing with time. The numerical results also show that the proposed sequential detector is very powerful in detecting weak objects, especially compared with the optimal FSS detection procedure based on the LRT.

This chapter is organized as follows. The test statistic of the joint object detection and system state estimation algorithm based on the Kalman filter (KF) is derived in Section 2.2. In Section 2.3, the new joint detection and system state estimation algorithm by using Wald's SPRT is proposed. To evaluate the performance of proposed algorithm, the first and second moments of test statistic are derived under both hypotheses. In Section 2.4, the performance of proposed algorithms is evaluated via simulations. Finally, this chapter is concluded in Section 2.5.

2.2 Problem Formulation

Let us assume that under hypothesis H_1, an object exists and its motion could be modeled by the following discrete-time linear system state equation [4]

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{v}_k \tag{2.1}$$

where $\mathbf{x}_k$ is $n_x \times 1$ state vector at time k, $\mathbf{F}_k$ is $n_x \times n_x$ state transition matrix at time k, and $\{\mathbf{v}_k\}$ is a sequence of white Gaussian process noise with $\mathrm{E}(\mathbf{v}_k) = \mathbf{0}$ and $\mathrm{E}(\mathbf{v}_k \mathbf{v}_k^T) = \mathbf{Q}_k$ for all $k = 0, 1, 2, \cdots$.

The measurement equation is

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \tag{2.2}$$

where $\mathbf{z}_k$ is the $n_z \times 1$ measurement vector at time k, $\mathbf{H}_k$ is the $n_z \times n_x$ measurement matrix at time k, and $\mathbf{w}_k$ is the measurement noise at time k. Also, $\{\mathbf{w}_k\}$ is a sequence of white Gaussian measurement noise with $\mathrm{E}(\mathbf{w}_k) = \mathbf{0}$ and $\mathrm{E}(\mathbf{w}_k \mathbf{w}_k^T) = \mathbf{R}_{w_k}$ for $k = 1, 2, \cdots$.

Let us assume that under hypothesis H_0, no object exists and the measurement is purely noise

$$\mathbf{z}_k = \mathbf{u}_k \tag{2.3}$$

where $\mathbf{u}_k$s are independent and follow Gaussian distribution with mean $\boldsymbol{\mu}_k$ and covariance $\mathbf{R}_{u_k}$. $\mathbf{x}_0$, $\mathbf{v}_k$, $\mathbf{w}_k$, and $\mathbf{u}_k$ are assumed to be independent with each other for all k.

To solve the joint detection and system state estimation problem, it is important to derive the log-likelihood ratio for the measurements accumulated up to the Kth step, which is

$$\log\Lambda(\mathbf{z}_{1:K}) = \log\frac{p(\mathbf{z}_{1:K}|H_1)}{p(\mathbf{z}_{1:K}|H_0)} \tag{2.4}$$

Under hypothesis H_1, the likelihood function $p(\mathbf{z}_{1:K}|H_1)$ can be calculated using chain rule as follows

$$p(\mathbf{z}_{1:K}|H_1) = p(\mathbf{z}_1|H_1) \prod_{k=1}^{K-1} p(\mathbf{z}_{k+1}|\mathbf{z}_{1:k}, H_1) \tag{2.5}$$

Since the measurements are independent over time under hypothesis H_0 and substituting (2.5) in (2.4), the log-likelihood ratio can be written in the following summation form

$$\begin{aligned} \log\Lambda(\mathbf{z}_{1:K}) &= \log\frac{p(\mathbf{z}_1|H_1)\prod_{k=1}^{K-1} p(\mathbf{z}_{k+1}|\mathbf{z}_{1:k}, H_1)}{\prod_{k=1}^{K} p(\mathbf{z}_k|H_0)} \\ &= \sum_{k=1}^{K} \log\frac{p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, H_1)}{p(\mathbf{z}_k|H_0)} = \sum_{k=1}^{K} \Theta_k \end{aligned} \tag{2.6}$$

in which

$$\Theta_k \triangleq \log\frac{p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, H_1)}{p(\mathbf{z}_k|H_0)} \tag{2.7}$$

and $\mathbf{z}_{1:0}$ is an empty set.

By using the Kalman filter in this joint detection and system state estimation problem in a manner similar to using the EKF in [30], we obtain $p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, H_1)$ as follows

$$\begin{aligned} &p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, H_1) \\ &= |2\pi\mathbf{S}_k|^{-\frac{1}{2}} e^{-\frac{(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T \mathbf{S}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})}{2}} \end{aligned} \tag{2.8}$$

where $\mathbf{S}_k$ is the measurement residue covariance, and $\hat{\mathbf{x}}_{k|k-1}$ is the predicted state given the accumulated measurements $\mathbf{z}_{1:k-1}$. Under hypothesis H_0, it is easy to show that

$$p(\mathbf{z}_k|H_0) = |2\pi\mathbf{R}_{u_k}|^{-\frac{1}{2}} e^{-\frac{(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k)}{2}} \tag{2.9}$$

Substituting (2.8) and (2.9) in (2.7), we have

$$\begin{aligned}\Theta_k =& \frac{1}{2}\log\frac{|\mathbf{R}_{u_k}|}{|\mathbf{S}_k|} + \frac{1}{2}(\mathbf{z}_k - \boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k) \\ & - \frac{1}{2}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T\mathbf{S}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})\end{aligned} \tag{2.10}$$

Let $t_{\mathrm{I}}(\mathbf{z}_{1:K}) \triangleq 2\sum_{k=1}^{K}\Theta_k$ be the hypothesis testing statistic. According to (2.6) and (2.10), we have

$$\begin{aligned}& t_{\mathrm{I}}(\mathbf{z}_{1:K}) \\ &= 2\log\Lambda(\mathbf{z}_{1:K}) \\ &= \sum_{k=1}^{K}\left[\log\frac{|\mathbf{R}_{u_k}|}{|\mathbf{S}_k|} + (\mathbf{z}_k - \boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k)\right. \\ &\quad \left. - \left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)^T\mathbf{S}_k^{-1}\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)\right]\end{aligned} \tag{2.11}$$

2.3 Joint Sequential Detection and System State Estimation

2.3.1 Applying Wald's SPRT

According to [15], the Wald's SPRT is weakly admissible even if the samples are dependent. A sequential test for deciding between H_0 and H_1, whose stopping variable is N_s and whose error probabilities are α and β, will be called inadmissible if there exists an alternative test T', with stopping variable N_s' and error probabilities α' and β', such that

$$\alpha' \leqslant \alpha,\ P(N_s' > K|H_0) \leqslant P(N_s > K|H_0),$$

$$\beta' \leqslant \beta,\ P(N_s' > K|H_1) \leqslant P(N_s > K|H_1),\ K = 1, 2, \ldots$$

with at least one of these inequalities strict. T is admissible if no such T' exists. A test T will be called weakly admissible if there is no alternative test T' which not only improves upon T in the sense of admissibility but also stops strictly no later than T

does under H_0 and H_1 [15, 17].

Since the ASN required by Wald's SPRT is smaller than the number of samples required by the FSS test [13] and Wald's SPRT is weakly admissible when samples are dependent [15], we propose a new joint sequential detection and system state estimation algorithm by using Wald's SPRT. Assume that Wald's SPRT procedure will continue by taking an additional measurement if

$$B < \Lambda(\mathbf{z}_{1:K}) < A \tag{2.12}$$

For purposes of practical computation, let us take the logarithm of (2.12) as follows

$$\log B < \log\Lambda(\mathbf{z}_{1:K}) < \log A \tag{2.13}$$

Substituting (2.11) in (2.13) we have

$$2\log B < t_{\mathrm{I}}(\mathbf{z}_{1:K}) < 2\log A \tag{2.14}$$

Hypothesis H_1 will be accepted and the sequential test will terminate as long as

$$t_{\mathrm{I}}(\mathbf{z}_{1:K}) \geqslant 2\log A \tag{2.15}$$

Hypothesis H_0 will be accepted and the sequential test will terminate as long as

$$t_{\mathrm{I}}(\mathbf{z}_{1:K}) \leqslant 2\log B \tag{2.16}$$

2.3.2 Expected Values of the Test Statistic

To gain some insights on the termination of the proposed algorithm, we derive the expectation of test statistic $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ under both hypotheses. In the derivation, we assume that $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$, where $\hat{\mathbf{x}}_{0|0}$ and $\mathbf{P}_{0|0}$ are the mean and covariance matrix associated with the prior PDF of the object state at time 0.

2.3.2.1 Expectation of Test Statistic $t_{\mathbf{I}}(\mathbf{z}_{1:K})$ Under H_1

Under hypothesis H_1, based on (2.11), we have

$$\begin{aligned}
&\mathrm{E}\left[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1\right] \\
&= \sum_{k=1}^{K} \left\{ \log \frac{|\mathbf{R}_{u_k}|}{|\mathbf{S}_k|} + \mathrm{E}\left[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k) \,|H_1 \right] \right. \\
&\quad \left. - \mathrm{E}\left[\left(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right)^T \mathbf{S}_k^{-1} \left(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}\right) |H_1 \right] \right\}
\end{aligned} \tag{2.17}$$

In this subsection, for simplicity and without ambiguity, we skip the conditioning on hypothesis H_1 in the notations.

Let us consider the second term $\mathrm{E}[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k)]$ in (2.17), which can be expanded as

$$\begin{aligned}
&\mathrm{E}[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k)] \\
&= \mathrm{tr}[\mathbf{R}_{u_k}^{-1} \mathrm{Var}(\mathbf{z}_k - \boldsymbol{\mu}_k)] + \mathrm{E}(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} \mathrm{E}(\mathbf{z}_k - \boldsymbol{\mu}_k)
\end{aligned} \tag{2.18}$$

where $\mathrm{Var}(\mathbf{y}) = \mathrm{E}\left[(\mathbf{y} - \mathrm{E}(\mathbf{y}))(\mathbf{y} - \mathrm{E}(\mathbf{y}))^T\right]$ represents the covariance matrix of a random vector $\mathbf{y}$.

Clearly, to calculate $\mathrm{E}[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k)]$, we need both the expectation and the covariance matrix of $\mathbf{z}_k - \boldsymbol{\mu}_k$. Since $\mathrm{E}(\mathbf{w}_k) = \mathrm{E}(\mathbf{v}_k) = 0$ for all k, the expectation of $\mathbf{z}_k - \boldsymbol{\mu}_k$ is

$$\begin{aligned}
\mathrm{E}(\mathbf{z}_k - \boldsymbol{\mu}_k) &= \mathrm{E}(\mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k) - \boldsymbol{\mu}_k \\
&= \mathbf{H}_k \mathrm{E}(\mathbf{F}_{k-1} \mathbf{x}_{k-1} + \mathbf{v}_{k-1}) - \boldsymbol{\mu}_k \\
&= \mathbf{H}_k \mathbf{F}_{k-1} \mathrm{E}(\mathbf{x}_{k-1}) - \boldsymbol{\mu}_k
\end{aligned} \tag{2.19}$$

We can repeat this process k times till we get the following results:

$$\begin{aligned} \mathrm{E}(\mathbf{z}_k - \boldsymbol{\mu}_k) &= \mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \, \mathrm{E}(\mathbf{x}_0) - \boldsymbol{\mu}_k \\ &= \mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \hat{\mathbf{x}}_{0|0} - \boldsymbol{\mu}_k \end{aligned} \tag{2.20}$$

where the product of the matrices is defined as

$$\prod_{i=i_1}^{i_2} \mathbf{A}_i = \mathbf{A}_{i_1} \mathbf{A}_{i_1+1} \cdots \mathbf{A}_{i_2} \tag{2.21}$$

Since $\boldsymbol{\mu}_k$ is deterministic, the covariance matrix of $\mathbf{z}_k - \boldsymbol{\mu}_k$ is the same as that of $\mathbf{z}_k$.

$$\begin{aligned} &\mathrm{Var}(\mathbf{z}_k - \boldsymbol{\mu}_k) = \mathrm{Var}(\mathbf{z}_k) = \mathrm{Var}(\mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k) \\ &= \mathbf{H}_k \, \mathrm{Var}(\mathbf{x}_k) \mathbf{H}_k^T + \mathbf{R}_{w_k} \\ &= \mathbf{H}_k \, \mathrm{Var}(\mathbf{F}_{k-1} \mathbf{x}_{k-1} + \mathbf{v}_{k-1}) \mathbf{H}_k^T + \mathbf{R}_{w_k} \\ &= \mathbf{H}_k \mathbf{F}_{k-1} \, \mathrm{Var}(\mathbf{x}_{k-1}) (\mathbf{H}_k \mathbf{F}_{k-1})^T + \mathbf{H}_k \mathbf{Q}_{k-1} \mathbf{H}_k^T + \mathbf{R}_{w_k} \end{aligned} \tag{2.22}$$

Repeating this process k times, we have

$$\begin{aligned} &\mathrm{Var}(\mathbf{z}_k - \boldsymbol{\mu}_k) = \mathrm{Var}(\mathbf{z}_k) \\ &= \mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \, \mathrm{Var}(\mathbf{x}_0) \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \right)^T + \mathbf{R}_{w_k} \\ &\quad + \sum_{i=0}^{k-1} \mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j} \mathbf{Q}_i \left(\mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j} \right)^T \\ &= \mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \mathbf{P}_{0|0} \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \right)^T + \mathbf{R}_{w_k} \\ &\quad + \sum_{i=0}^{k-1} \mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j} \mathbf{Q}_i \left(\mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j} \right)^T \end{aligned} \tag{2.23}$$

where $\prod_{j=1}^{0} \mathbf{F}_{k-j} = \mathbf{I}$.

Substituting (2.20) and (2.23) in (2.18), we finally have

$$\begin{aligned}
&\mathrm{E}[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k)] \\
&= \mathrm{tr}\left[\mathbf{R}_{u_k}^{-1} \mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \mathbf{P}_{0|0} \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \right)^T + \mathbf{R}_{u_k}^{-1} \mathbf{R}_{w_k} \right. \\
&\quad \left. + \mathbf{R}_{u_k}^{-1} \sum_{i=0}^{k-1} \mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j} \mathbf{Q}_i \left(\mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j} \right)^T \right] \\
&\quad + \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \hat{\mathbf{x}}_{0|0} - \boldsymbol{\mu}_k \right)^T \mathbf{R}_{u_k}^{-1} \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \hat{\mathbf{x}}_{0|0} - \boldsymbol{\mu}_k \right)
\end{aligned} \tag{2.24}$$

Now let us consider the third term in (2.17). It is known that in a Kalman filter, the innovation $\boldsymbol{\nu}_k \triangleq \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$ is a zero-mean Gaussian random variable with covariance matrix $\mathbf{S}_k$ [4]. Then, the normalized innovation squared $(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})^T \mathbf{S}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$ follows a Chi-square distribution with degree of freedom n_z, where n_z is the dimension of the measurement $\mathbf{z}_k$. Since the expected value of a Chi-square distributed random variable is its degree of freedom, we have

$$\mathrm{E}\left[(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})^T \mathbf{S}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \right] = n_z \tag{2.25}$$

Finally, by substituting (2.24) and (2.25) in (2.17), we can obtain $\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$ and summarize the result in the following proposition.

Proposition 1 *The expectation of test statistic $t_I(\mathbf{z}_{1:K})$ under hypothesis H_1 is pro-*

vided as follows

$$
\begin{aligned}
&\mathrm{E}[t_I(\mathbf{z}_{1:K})|H_1] \\
&= \sum_{k=1}^{K} \Bigg\{ log\frac{|\mathbf{R}_{u_k}|}{|\mathbf{S}_k|} - n_z + \mathrm{tr}(\mathbf{R}_{u_k}^{-1}\mathbf{R}_{w_k}) \\
&\quad + \mathrm{tr}\left[\mathbf{R}_{u_k}^{-1}\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i}\mathbf{P}_{0|0} \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i} \right)^T \right. \\
&\quad \left. + \mathbf{R}_{u_k}^{-1} \sum_{i=0}^{k-1} \mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j}\mathbf{Q}_i \left(\mathbf{H}_k \prod_{j=1}^{k-i-1} \mathbf{F}_{k-j} \right)^T \right] \\
&\quad + \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i}\hat{\mathbf{x}}_{0|0} - \boldsymbol{\mu}_k \right)^T \mathbf{R}_{u_k}^{-1} \left(\mathbf{H}_k \prod_{i=1}^{k} \mathbf{F}_{k-i}\hat{\mathbf{x}}_{0|0} - \boldsymbol{\mu}_k \right) \Bigg\}
\end{aligned}
\tag{2.26}
$$

2.3.2.2 Expectation of Test Statistic $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ Under H_0

Using (2.11), we have the expectation of test statistic $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ under H_0:

$$
\begin{aligned}
&\mathrm{E}\left[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0\right] \\
&= \sum_{k=1}^{K} \Bigg\{ \log \frac{|\mathbf{R}_{u_k}|}{|\mathbf{S}_k|} + \mathrm{E}\left[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k) \,|H_0 \right] \\
&\quad - \mathrm{E}\left[\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)^T \mathbf{S}_k^{-1} \left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right) |H_0 \right] \Bigg\}
\end{aligned}
\tag{2.27}
$$

In this subsection, without ambiguity and for simplicity, again we skip the conditioning on hypothesis H_0 in the notations.

Let us first consider the second term in (2.27). Since under hypothesis H_0, $(\mathbf{z}_k - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0},\ \mathbf{R}_{u_k})$, it is very easy to show that

$$
\mathrm{E}\left[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k) \right] = n_z \tag{2.28}
$$

Similar to the derivation in Subsection 2.3.2.1, to find the third term in (2.27), the expectation and covariance matrix of $\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$ need to be calculated. First,

let us investigate its expectation:

$$
\begin{aligned}
&\mathrm{E}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \\
&= \boldsymbol{\mu}_k - \mathbf{H}_k\mathbf{F}_{k-1}\,\mathrm{E}(\hat{\mathbf{x}}_{k-1|k-1}) \\
&= \boldsymbol{\mu}_k - \mathbf{H}_k\mathbf{F}_{k-1}\,\mathrm{E}\left[\hat{\mathbf{x}}_{k-1|k-2} + \mathbf{W}_{k-1}\left(\mathbf{z}_{k-1} - \mathbf{H}_{k-1}\hat{\mathbf{x}}_{k-1|k-2}\right)\right] \\
&= \boldsymbol{\mu}_k - \mathbf{H}_k\mathbf{F}_{k-1}\mathbf{W}_{k-1}\boldsymbol{\mu}_{k-1} \\
&\quad - \mathbf{H}_k\mathbf{F}_{k-1}(\mathbf{I} - \mathbf{W}_{k-1}\mathbf{H}_{k-1})\,\mathrm{E}(\hat{\mathbf{x}}_{k-1|k-2}) \\
&= \boldsymbol{\mu}_k - \mathbf{H}_k\mathbf{F}_{k-1}\mathbf{W}_{k-1}\boldsymbol{\mu}_{k-1} \\
&\quad - \mathbf{H}_k\mathbf{F}_{k-1}(\mathbf{I} - \mathbf{W}_{k-1}\mathbf{H}_{k-1})\mathbf{F}_{k-2}\,\mathrm{E}(\hat{\mathbf{x}}_{k-2|k-2})
\end{aligned}
\tag{2.29}
$$

where $\mathbf{W}_k$ is the Kalman filter gain at time k. Repeating this process, we have

$$
\begin{aligned}
&\mathrm{E}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \\
&= \boldsymbol{\mu}_k - \mathbf{H}_k\mathbf{F}_{k-1}\left\{\prod_{j=1}^{k-1}\left[(\mathbf{I} - \mathbf{W}_{k-j}\mathbf{H}_{k-j})\,\mathbf{F}_{k-j-1}\right]\hat{\mathbf{x}}_{0|0}\right. \\
&\quad \left. - \sum_{i=1}^{k-1}\prod_{j=1}^{i-1}\left[(\mathbf{I} - \mathbf{W}_{k-j}\mathbf{H}_{k-j})\,\mathbf{F}_{k-j-1}\right]\mathbf{W}_{k-i}\boldsymbol{\mu}_{k-i}\right\} \\
&= \boldsymbol{\mu}_k - \mathbf{B}_{k,k}\hat{\mathbf{x}}_{0|0} - \sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\boldsymbol{\mu}_{k-i}
\end{aligned}
\tag{2.30}
$$

in which we define $\prod_{j=1}^{0}\left[(\mathbf{I} - \mathbf{W}_{k-j}\mathbf{H}_{k-j})\,\mathbf{F}_{k-j-1}\right] = \mathbf{I}$, $\sum_{i=1}^{0}\mathbf{A}_i = \mathbf{0}$, and

$$
\mathbf{B}_{k,i} = \mathbf{H}_k\mathbf{F}_{k-1}\prod_{j=1}^{i-1}\left[(\mathbf{I} - \mathbf{W}_{k-j}\mathbf{H}_{k-j})\,\mathbf{F}_{k-j-1}\right] \tag{2.31}
$$

Since $\hat{\mathbf{x}}_{k|k-1}$ is a function of $\mathbf{z}_{1:k-1}$, it is independent of $\mathbf{z}_k$. Considering this fact,

the covariance matrix of $\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$ is

$$\begin{aligned}
&\mathrm{Var}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \\
=&\mathbf{H}_k\mathbf{F}_{k-1}\,\mathrm{Var}(\hat{\mathbf{x}}_{k-1|k-1})(\mathbf{H}_k\mathbf{F}_{k-1})^T + \mathbf{R}_{u_k} \\
=&\mathbf{H}_k\mathbf{F}_{k-1}\,\mathrm{Var}\left[(\mathbf{I} - \mathbf{W}_{k-1}\mathbf{H}_{k-1})\,\hat{\mathbf{x}}_{k-1|k-2}\right. \\
&\left.+\mathbf{W}_{k-1}\mathbf{z}_{k-1}\right](\mathbf{H}_k\mathbf{F}_{k-1})^T + \mathbf{R}_{u_k}
\end{aligned} \tag{2.32}$$

Repeating this process, we finally have

$$\begin{aligned}
&\mathrm{Var}(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}) \\
=&\mathbf{R}_{u_k} + \mathbf{H}_k\mathbf{F}_{k-1}\prod_{i=1}^{k-1}\left[(\mathbf{I} - \mathbf{W}_{k-i}\mathbf{H}_{k-i})\,\mathbf{F}_{k-i-1}\right]\mathrm{Var}(\hat{\mathbf{x}}_{0|0}) \\
&\cdot\left\{\mathbf{H}_k\mathbf{F}_{k-1}\prod_{i=1}^{k-1}\left[(\mathbf{I} - \mathbf{W}_{k-i}\mathbf{H}_{k-i})\,\mathbf{F}_{k-i-1}\right]\right\}^T \\
&+\sum_{i=1}^{k-1}\mathbf{H}_k\mathbf{F}_{k-1}\prod_{j=1}^{i-1}\left[(\mathbf{I} - \mathbf{W}_{k-j}\mathbf{H}_{k-j})\,\mathbf{F}_{k-j-1}\right]\mathbf{W}_{k-i}\mathbf{R}_{u_{k-i}} \\
&\cdot\left\{\mathbf{H}_k\mathbf{F}_{k-1}\prod_{j=1}^{i-1}\left[(\mathbf{I} - \mathbf{W}_{k-j}\mathbf{H}_{k-j})\,\mathbf{F}_{k-j-1}\right]\mathbf{W}_{k-i}\right\}^T \\
=&\mathbf{R}_{u_k} + \sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\mathbf{R}_{u_{k-i}}\,(\mathbf{B}_{k,i}\mathbf{W}_{k-i})^T
\end{aligned} \tag{2.33}$$

in which we use the fact that $\hat{\mathbf{x}}_{0|0}$ is a constant.

Similar to (2.18), the third term in (2.27) is expanded by using (2.30) and (2.33)

as follows

$$
\begin{aligned}
&\mathrm{E}\left[\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)^T \mathbf{S}_k^{-1}\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)\right] \\
&= \mathrm{tr}\left[\mathbf{S}_k^{-1}\,\mathrm{Var}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})\right] \\
&\quad + \mathrm{E}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T\mathbf{S}_k^{-1}\,\mathrm{E}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \\
&= \mathrm{tr}\left[\mathbf{S}_k^{-1}\mathbf{R}_{u_k} + \mathbf{S}_k^{-1}\sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\mathbf{R}_{u_{k-i}}(\mathbf{B}_{k,i}\mathbf{W}_{k-i})^T\right] \\
&\quad + \left(\boldsymbol{\mu}_k - \mathbf{B}_{k,k}\hat{\mathbf{x}}_{0|0} - \sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\boldsymbol{\mu}_{k-i}\right)^T \mathbf{S}_k^{-1} \\
&\qquad \cdot \left(\boldsymbol{\mu}_k - \mathbf{B}_{k,k}\hat{\mathbf{x}}_{0|0} - \sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\boldsymbol{\mu}_{k-i}\right)
\end{aligned}
\tag{2.34}
$$

Finally, substituting (2.28) and (2.34) in (2.27), we have the following proposition.

Proposition 2 *The expectation of test statistic $t_I(\mathbf{z}_{1:K})$ under H_0 when using the Kalman filter is*

$$
\begin{aligned}
&\mathrm{E}\left[t_I(\mathbf{z}_{1:K})|H_0\right] \\
&= \sum_{k=1}^{K}\Bigg\{ log\frac{|\mathbf{R}_{u_k}|}{|\mathbf{S}_k|} + n_z \\
&\quad - \mathrm{tr}\left[\mathbf{S}_k^{-1}\mathbf{R}_{u_k} + \mathbf{S}_k^{-1}\sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\mathbf{R}_{u_{k-i}}(\mathbf{B}_{k,i}\mathbf{W}_{k-i})^T\right] \\
&\quad - \left(\boldsymbol{\mu}_k - \mathbf{B}_{k,k}\hat{\mathbf{x}}_{0|0} - \sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\boldsymbol{\mu}_{k-i}\right)^T \mathbf{S}_k^{-1} \\
&\qquad \cdot \left(\boldsymbol{\mu}_k - \mathbf{B}_{k,k}\hat{\mathbf{x}}_{0|0} - \sum_{i=1}^{k-1}\mathbf{B}_{k,i}\mathbf{W}_{k-i}\boldsymbol{\mu}_{k-i}\right)\Bigg\}
\end{aligned}
\tag{2.35}
$$

According to [31], the Kullback-Leibler distance between $p(\mathbf{z}_{1:K}|H_1)$ and $p(\mathbf{z}_{1:K}|H_0)$

has the following relationship with $\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$.

$$\begin{aligned} & D(p(\mathbf{z}_{1:K}|H_1)||p(\mathbf{z}_{1:K}|H_0)) \\ &= \mathrm{E}_{p(\mathbf{z}_{1:K}|H_1)}\left[\log\frac{p(\mathbf{z}_{1:K}|H_1)}{p(\mathbf{z}_{1:K}|H_0)}\right] \\ &= \mathrm{E}_{p(\mathbf{z}_{1:K}|H_1)}[\log\Lambda(\mathbf{z}_{1:K})] \\ &= \frac{1}{2}\,\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1] \end{aligned} \tag{2.36}$$

The Kullback-Leibler distance between $p(\mathbf{z}_{1:K}|H_0)$ and $p(\mathbf{z}_{1:K}|H_1)$ has the following relationship with $\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0]$.

$$\begin{aligned} & D(p(\mathbf{z}_{1:K}|H_0)||p(\mathbf{z}_{1:K}|H_1)) \\ &= \mathrm{E}_{p(\mathbf{z}_{1:K}|H_0)}\left[\log\frac{p(\mathbf{z}_{1:K}|H_0)}{p(\mathbf{z}_{1:K}|H_1)}\right] \\ &= -\,\mathrm{E}_{p(\mathbf{z}_{1:K}|H_0)}[\log\Lambda(\mathbf{z}_{1:K})] \\ &= -\frac{1}{2}\,\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0] \end{aligned} \tag{2.37}$$

2.3.3 Variances of the Test Statistic

The variance of test statistic $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ in (2.11) does not depend on $\log\frac{|\mathbf{R}_{u_k}|}{|\mathbf{S}_k|}$ because this term is deterministic. So, it contains two parts. Part I contains all the variances of the terms in sequence $(\mathbf{z}_k-\boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}(\mathbf{z}_k-\boldsymbol{\mu}_k)$ where $k=1,2,...,K$ and sequence $-(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T\mathbf{S}_k^{-1}(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})$ where $k=1,2,...,K$. Part II contains all the covariances of the terms in the two sequences. It can be summarized as follows

$$
\begin{aligned}
\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})] &= \sum_{k=1}^{K} \mathrm{Var}\left[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k)\right] \\
&+ \sum_{k=1}^{K} \mathrm{Var}\left[(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})^T \mathbf{S}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})\right] \\
&+ \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \mathrm{Cov}[(\mathbf{z}_i - \boldsymbol{\mu}_i)^T \mathbf{R}_{u_i}^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_i), \\
&\quad (\mathbf{z}_j - \boldsymbol{\mu}_j)^T \mathbf{R}_{u_j}^{-1} (\mathbf{z}_j - \boldsymbol{\mu}_j)] \\
&+ \sum_{i=1}^{K} \sum_{\substack{j=1 \\ j \neq i}}^{K} \mathrm{Cov}[(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1})^T \mathbf{S}_i^{-1} (\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}), \\
&\quad (\mathbf{z}_j - \mathbf{H}_j \hat{\mathbf{x}}_{j|j-1})^T \mathbf{S}_j^{-1} (\mathbf{z}_j - \mathbf{H}_j \hat{\mathbf{x}}_{j|j-1})] \\
&- 2 \sum_{i=1}^{K} \sum_{j=1}^{K} \mathrm{Cov}[(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1})^T \mathbf{S}_i^{-1} (\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}), \\
&\quad (\mathbf{z}_j - \boldsymbol{\mu}_j)^T \mathbf{R}_{u_j}^{-1} (\mathbf{z}_j - \boldsymbol{\mu}_j)]
\end{aligned}
\tag{2.38}
$$

where $\mathrm{Cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}(\mathbf{x}))(\mathbf{y} - \mathrm{E}(\mathbf{y}))^T\right]$ denotes the covariance between $\mathbf{x}$ and $\mathbf{y}$.

To derive part I, we need the expression of $\mathrm{Cov}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}, \mathbf{x}^T\mathbf{B}\mathbf{x}\right)$ which denotes the covariance between $\mathbf{x}^T\mathbf{A}\mathbf{x}$ and $\mathbf{x}^T\mathbf{B}\mathbf{x}$. Similarly, $\mathrm{Cov}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}, \mathbf{y}^T\mathbf{B}\mathbf{y}\right)$ is the key to derive part II.

Assuming $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices and $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu},\ \mathbf{P})$, the covariance between $\mathbf{x}^T\mathbf{A}\mathbf{x}$ and $\mathbf{x}^T\mathbf{B}\mathbf{x}$ is [32]

$$
\mathrm{Cov}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}, \mathbf{x}^T\mathbf{B}\mathbf{x}\right) = 2\,\mathrm{tr}(\mathbf{APBP}) + 4\boldsymbol{\mu}^T\mathbf{APB}\boldsymbol{\mu} \tag{2.39}
$$

To derive the covariance between $\mathbf{x}^T\mathbf{A}\mathbf{x}$ and $\mathbf{y}^T\mathbf{B}\mathbf{y}$, we need to derive $\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{y}^T\mathbf{B}\mathbf{y})$ first. To reduce the complexity of the derivation of $\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{y}^T\mathbf{B}\mathbf{y})$, we introduce a

vector and a matrix

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are symmetric matrices.

Let us denote the mean and covariance matrix of $\mathbf{x}$ as $\boldsymbol{\mu}_\mathbf{x}$ and $\mathbf{P}_\mathbf{x}$ respectively, and the mean and covariance matrix of $\mathbf{y}$ as $\boldsymbol{\mu}_\mathbf{y}$ and $\mathbf{P}_\mathbf{y}$ respectively. The covariance $\mathrm{Cov}(\mathbf{x}, \mathbf{y})$ is denoted by $\mathbf{P}_\mathbf{xy}$. Then, the mean and covariance matrix of $\mathbf{z}$ are as follows

$$\boldsymbol{\mu}_\mathbf{z} = \begin{bmatrix} \boldsymbol{\mu}_\mathbf{x} \\ \boldsymbol{\mu}_\mathbf{y} \end{bmatrix} \quad \text{and} \quad \mathbf{P}_\mathbf{z} = \begin{bmatrix} \mathbf{P}_\mathbf{x} & \mathbf{P}_\mathbf{xy} \\ \mathbf{P}_\mathbf{yx} & \mathbf{P}_\mathbf{y} \end{bmatrix} \tag{2.40}$$

The relationship between $\mathbf{x}^T\mathbf{A}\mathbf{x}$, $\mathbf{y}^T\mathbf{B}\mathbf{y}$ and $\mathbf{z}^T\mathbf{C}\mathbf{z}$ is as follows

$$\begin{aligned} \mathbf{z}^T\mathbf{C}\mathbf{z} &= \begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \\ &= \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{y}^T\mathbf{B}\mathbf{y} \end{aligned} \tag{2.41}$$

Then, we have

$$\begin{aligned} \left(\mathbf{z}^T\mathbf{C}\mathbf{z}\right)^2 &= \left(\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{y}^T\mathbf{B}\mathbf{y}\right)^2 \\ &= \left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)^2 + \left(\mathbf{y}^T\mathbf{B}\mathbf{y}\right)^2 + 2\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{y}^T\mathbf{B}\mathbf{y} \end{aligned} \tag{2.42}$$

Taking expectation on both sides of (2.42), we have

$$\begin{aligned} &\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{y}^T\mathbf{B}\mathbf{y}) \\ =&\frac{1}{2}\left\{\mathrm{E}\left[\left(\mathbf{z}^T\mathbf{C}\mathbf{z}\right)^2\right] - \mathrm{E}\left[\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)^2\right] - \mathrm{E}\left[\left(\mathbf{y}^T\mathbf{B}\mathbf{y}\right)^2\right]\right\} \end{aligned} \tag{2.43}$$

Now we need to derive the expression of $\mathrm{E}\left[\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)^2\right]$. According to (2.39), we have

$$\mathrm{Var}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = 2\,\mathrm{tr}\left[(\mathbf{A}\mathbf{P}_\mathbf{x})^2\right] + 4\boldsymbol{\mu}_\mathbf{x}^T\mathbf{A}\mathbf{P}_\mathbf{x}\mathbf{A}\boldsymbol{\mu}_\mathbf{x} \tag{2.44}$$

We also know that $\mathrm{E}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)$ [32]

$$\mathrm{E}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = \mathrm{tr}(\mathbf{A}\mathbf{P}_\mathbf{x}) + \boldsymbol{\mu}_\mathbf{x}^T\mathbf{A}\boldsymbol{\mu}_\mathbf{x} \tag{2.45}$$

Then, $\mathrm{E}\left[\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)^2\right]$ is

$$\begin{aligned}
&\mathrm{E}\left[\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)^2\right] \\
&= \mathrm{Var}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right) + \left[\mathrm{E}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)\right]^2 \\
&= \left(\boldsymbol{\mu}_\mathbf{x}^T\mathbf{A}\boldsymbol{\mu}_\mathbf{x}\right)^2 + 2\,\mathrm{tr}(\mathbf{A}\mathbf{P}_\mathbf{x})\boldsymbol{\mu}_\mathbf{x}^T\mathbf{A}\boldsymbol{\mu}_\mathbf{x} + [\mathrm{tr}(\mathbf{A}\mathbf{P}_\mathbf{x})]^2 \\
&\quad + 2\,\mathrm{tr}\left[(\mathbf{A}\mathbf{P}_\mathbf{x})^2\right] + 4\boldsymbol{\mu}_\mathbf{x}^T\mathbf{A}\mathbf{P}_\mathbf{x}\mathbf{A}\boldsymbol{\mu}_\mathbf{x}
\end{aligned} \tag{2.46}$$

Similarly, $\mathrm{E}\left[\left(\mathbf{y}^T\mathbf{B}\mathbf{y}\right)^2\right]$ is

$$\begin{aligned}
&\mathrm{E}\left[\left(\mathbf{y}^T\mathbf{B}\mathbf{y}\right)^2\right] \\
&= \left(\boldsymbol{\mu}_\mathbf{y}^T\mathbf{B}\boldsymbol{\mu}_\mathbf{y}\right)^2 + 2\,\mathrm{tr}(\mathbf{B}\mathbf{P}_\mathbf{y})\boldsymbol{\mu}_\mathbf{y}^T\mathbf{B}\boldsymbol{\mu}_\mathbf{y} + [\mathrm{tr}(\mathbf{B}\mathbf{P}_\mathbf{y})]^2 \\
&\quad + 2\,\mathrm{tr}\left[(\mathbf{B}\mathbf{P}_\mathbf{y})^2\right] + 4\boldsymbol{\mu}_\mathbf{y}^T\mathbf{B}\mathbf{P}_\mathbf{y}\mathbf{B}\boldsymbol{\mu}_\mathbf{y}
\end{aligned} \tag{2.47}$$

$\mathrm{E}\left[\left(\mathbf{z}^T\mathbf{C}\mathbf{z}\right)^2\right]$ is derived via (2.40) as follows

$$\begin{aligned}
&\mathrm{E}\left[\left(\mathbf{z}^T\mathbf{C}\mathbf{z}\right)^2\right] \\
&= \left(\boldsymbol{\mu}_{\mathbf{z}}^T\mathbf{C}\boldsymbol{\mu}_{\mathbf{z}}\right)^2 + 2\,\mathrm{tr}(\mathbf{C}\mathbf{P}_{\mathbf{z}})\boldsymbol{\mu}_{\mathbf{z}}^T\mathbf{C}\boldsymbol{\mu}_{\mathbf{z}} + [\mathrm{tr}(\mathbf{C}\mathbf{P}_{\mathbf{z}})]^2 \\
&\quad + 2\,\mathrm{tr}\left[(\mathbf{C}\mathbf{P}_{\mathbf{z}})^2\right] + 4\boldsymbol{\mu}_{\mathbf{z}}^T\mathbf{C}\mathbf{P}_{\mathbf{z}}\mathbf{C}\boldsymbol{\mu}_{\mathbf{z}} \\
&= \left(\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{y}}^T\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}}\right)^2 \\
&\quad + 2\left[\mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}}) + \mathrm{tr}(\mathbf{B}\mathbf{P}_{\mathbf{y}})\right]\left(\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{y}}^T\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}}\right) \\
&\quad + [\mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}}) + \mathrm{tr}(\mathbf{B}\mathbf{P}_{\mathbf{y}})]^2 \\
&\quad + 2\left\{\mathrm{tr}\left[(\mathbf{A}\mathbf{P}_{\mathbf{x}})^2\right] + 2\,\mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}\mathbf{y}}\mathbf{B}\mathbf{P}_{\mathbf{y}\mathbf{x}}) + \mathrm{tr}\left[(\mathbf{B}\mathbf{P}_{\mathbf{y}})^2\right]\right\} \\
&\quad + 4\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}}^T & \boldsymbol{\mu}_{\mathbf{y}}^T \end{bmatrix}\begin{bmatrix} \mathbf{A}\mathbf{P}_{\mathbf{x}}\mathbf{A} & \mathbf{A}\mathbf{P}_{\mathbf{x}\mathbf{y}}\mathbf{B} \\ \mathbf{B}\mathbf{P}_{\mathbf{y}\mathbf{x}}\mathbf{A} & \mathbf{B}\mathbf{P}_{\mathbf{y}}\mathbf{B} \end{bmatrix}\begin{bmatrix} \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{bmatrix}
\end{aligned} \tag{2.48}$$

Substituting (2.46), (2.47) and (2.48) in (2.43), the mean of $\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{y}^T\mathbf{B}\mathbf{y}$ is derived as follows

$$\begin{aligned}
&\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{y}^T\mathbf{B}\mathbf{y}) \\
&= 2\,\mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}\mathbf{y}}\mathbf{B}\mathbf{P}_{\mathbf{y}\mathbf{x}}) + 4\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\mathbf{P}_{\mathbf{x}\mathbf{y}}\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}} \\
&\quad + \mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}})\,\mathrm{tr}(\mathbf{B}\mathbf{P}_{\mathbf{y}}) + \boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{y}}^T\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}} \\
&\quad + \mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}})\boldsymbol{\mu}_{\mathbf{y}}^T\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}} + \mathrm{tr}(\mathbf{B}\mathbf{P}_{\mathbf{y}})\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}
\end{aligned} \tag{2.49}$$

Since the product $\mathrm{E}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)\mathrm{E}\left(\mathbf{y}^T\mathbf{B}\mathbf{y}\right)$ is

$$\begin{aligned}
&\mathrm{E}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)\mathrm{E}\left(\mathbf{y}^T\mathbf{B}\mathbf{y}\right) \\
&= \left[\mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}}) + \boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}\right]\left[\mathrm{tr}(\mathbf{B}\mathbf{P}_{\mathbf{y}}) + \boldsymbol{\mu}_{\mathbf{y}}^T\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}}\right] \\
&= \mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}})\,\mathrm{tr}(\mathbf{B}\mathbf{P}_{\mathbf{y}}) + \boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}\boldsymbol{\mu}_{\mathbf{y}}^T\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}} \\
&\quad + \mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{x}})\boldsymbol{\mu}_{\mathbf{y}}^T\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}} + \mathrm{tr}(\mathbf{B}\mathbf{P}_{\mathbf{y}})\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\boldsymbol{\mu}_{\mathbf{x}}
\end{aligned} \tag{2.50}$$

the covariance between $\mathbf{x}^T\mathbf{A}\mathbf{x}$ and $\mathbf{y}^T\mathbf{B}\mathbf{y}$ is

$$\begin{aligned}
&\mathrm{Cov}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}, \mathbf{y}^T\mathbf{B}\mathbf{y}\right) \\
&= \mathrm{E}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{y}^T\mathbf{B}\mathbf{y}\right) - \mathrm{E}\left(\mathbf{x}^T\mathbf{A}\mathbf{x}\right)\mathrm{E}\left(\mathbf{y}^T\mathbf{B}\mathbf{y}\right) \\
&= 2\,\mathrm{tr}(\mathbf{A}\mathbf{P}_{\mathbf{xy}}\mathbf{B}\mathbf{P}_{\mathbf{yx}}) + 4\boldsymbol{\mu}_{\mathbf{x}}^T\mathbf{A}\mathbf{P}_{\mathbf{xy}}\mathbf{B}\boldsymbol{\mu}_{\mathbf{y}}
\end{aligned} \tag{2.51}$$

Note that (2.51) is a general case of (2.39).

2.3.3.1 Variance of Test Statistic $t_{\mathbf{I}}(\mathbf{z}_{1:K})$ Under H_1

$\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$ will be derived by substituting means, variances and covariances of $\mathbf{z}_k - \boldsymbol{\mu}_k$ and $\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$ in (2.39) and (2.51). In this subsection, for simplicity and without ambiguity, we skip the conditioning on hypothesis H_1 in the notations.

The five terms in (2.38) are derived separately in order. Replacing $\mathbf{A}$ and $\mathbf{B}$ by $\mathbf{R}_{u_k}^{-1}$ and $\mathbf{x}$ by $\mathbf{z}_k - \boldsymbol{\mu}_k$ in (2.39), the variance of $(\mathbf{z}_k - \boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k)$ is derived as follows

$$\begin{aligned}
&\mathrm{Var}\left[(\mathbf{z}_k - \boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_k)\right] \\
&= 2\,\mathrm{tr}\left(\mathbf{R}_{u_k}^{-1}\,\mathrm{Var}(\mathbf{z}_k - \boldsymbol{\mu}_k)\mathbf{R}_{u_k}^{-1}\,\mathrm{Var}(\mathbf{z}_k - \boldsymbol{\mu}_k)\right) \\
&\quad + 4\,\mathrm{E}(\mathbf{z}_k - \boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}\,\mathrm{Var}(\mathbf{z}_k - \boldsymbol{\mu}_k)\mathbf{R}_{u_k}^{-1}\,\mathrm{E}(\mathbf{z}_k - \boldsymbol{\mu}_k)
\end{aligned} \tag{2.52}$$

where $\mathrm{E}(\mathbf{z}_k - \boldsymbol{\mu}_k)$ and $\mathrm{Var}(\mathbf{z}_k - \boldsymbol{\mu}_k)$ have been derived as in (2.20) and (2.23), respectively.

Since $(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T\mathbf{S}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})$ follows Chi-square distribution with degree of freedom n_z, its variance is

$$\mathrm{Var}\left[(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T\mathbf{S}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})\right] = 2n_z \tag{2.53}$$

Now, we derive the covariance between $(\mathbf{z}_i - \boldsymbol{\mu}_i)^T\mathbf{R}_{u_i}^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_i)$ and $(\mathbf{z}_j - \boldsymbol{\mu}_j)^T\mathbf{R}_{u_j}^{-1}(\mathbf{z}_j -$

$\boldsymbol{\mu}_j$) via (2.51) as follows

$$\begin{aligned}
&\mathrm{Cov}[(\mathbf{z}_i-\boldsymbol{\mu}_i)^T\mathbf{R}_{u_i}^{-1}(\mathbf{z}_i-\boldsymbol{\mu}_i),(\mathbf{z}_j-\boldsymbol{\mu}_j)^T\mathbf{R}_{u_j}^{-1}(\mathbf{z}_j-\boldsymbol{\mu}_j)]\\
&=2\,\mathrm{tr}(\mathbf{R}_{u_i}^{-1}\mathbf{P}_{\mu 1ij}\mathbf{R}_{u_j}^{-1}\mathbf{P}_{\mu 1ij}^T)\\
&\quad+4\,\mathrm{E}(\mathbf{z}_i-\boldsymbol{\mu}_i)^T\mathbf{R}_{u_i}^{-1}\mathbf{P}_{\mu 1ij}\mathbf{R}_{u_j}^{-1}\,\mathrm{E}(\mathbf{z}_j-\boldsymbol{\mu}_j)
\end{aligned}\tag{2.54}$$

where $\mathbf{P}_{\mu 1ij}=\mathrm{Cov}\left(\mathbf{z}_i-\boldsymbol{\mu}_i,\mathbf{z}_j-\boldsymbol{\mu}_j\right)=\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)$ since $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are deterministic.

Now, we need to find $\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)$. When $i<j$, $\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)$ is

$$\begin{aligned}
\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)&=\mathrm{Cov}\left(\mathbf{H}_i\mathbf{x}_i+\mathbf{w}_i,\mathbf{H}_j\mathbf{x}_j+\mathbf{w}_j\right)\\
&=\mathbf{H}_i\,\mathrm{Cov}\left(\mathbf{x}_i,\mathbf{x}_j\right)\mathbf{H}_j^T\\
&=\mathbf{H}_i\,\mathrm{Cov}\left(\mathbf{x}_i,\mathbf{F}_{j-1}\mathbf{x}_{j-1}+\mathbf{v}_{j-1}\right)\mathbf{H}_j^T\\
&=\mathbf{H}_i\,\mathrm{Cov}\left(\mathbf{x}_i,\mathbf{x}_{j-1}\right)\left(\mathbf{H}_j\mathbf{F}_{j-1}\right)^T
\end{aligned}\tag{2.55}$$

Repeating this process, we have

$$\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)=\mathbf{H}_i\,\mathrm{Var}(\mathbf{x}_i)\left(\mathbf{H}_j\prod_{m=1}^{j-i}\mathbf{F}_{j-m}\right)^T\tag{2.56}$$

Similar to (2.23), $\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)$ when $i<j$ is

$$\begin{aligned}
&\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)\\
&=\mathbf{H}_i\prod_{m=1}^{i}\mathbf{F}_{i-m}\mathbf{P}_{0|0}\left(\mathbf{H}_j\prod_{m=1}^{j}\mathbf{F}_{j-m}\right)^T\\
&\quad+\sum_{m=0}^{i-1}\mathbf{H}_i\prod_{n=1}^{i-m-1}\mathbf{F}_{i-n}\mathbf{Q}_m\left(\mathbf{H}_j\prod_{n=1}^{j-m-1}\mathbf{F}_{j-n}\right)^T
\end{aligned}\tag{2.57}$$

Similarly, $\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)$ can be derived when $i>j$. In summary, $\mathrm{Cov}\left(\mathbf{z}_i,\mathbf{z}_j\right)$ is as

follows when $i \neq j$

$$\begin{aligned}
&\mathrm{Cov}\left(\mathbf{z}_i, \mathbf{z}_j\right) \\
&= \mathbf{H}_i \prod_{m=1}^{i} \mathbf{F}_{i-m} \mathbf{P}_{0|0} \left(\mathbf{H}_j \prod_{m=1}^{j} \mathbf{F}_{j-m} \right)^T \\
&\quad + \sum_{m=0}^{\min(i,j)-1} \mathbf{H}_i \prod_{n=1}^{i-m-1} \mathbf{F}_{i-n} \mathbf{Q}_m \left(\mathbf{H}_j \prod_{n=1}^{j-m-1} \mathbf{F}_{j-n} \right)^T
\end{aligned} \tag{2.58}$$

Then, the covariance between $(\mathbf{z}_i - \boldsymbol{\mu}_i)^T \mathbf{R}_{u_i}^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_i)$ and $(\mathbf{z}_j - \boldsymbol{\mu}_j)^T \mathbf{R}_{u_j}^{-1} (\mathbf{z}_j - \boldsymbol{\mu}_j)$ can be got by substituting (2.20) and (2.58) in (2.54).

The fourth term in (2.38) is the covariance between two different terms in the sequence $-(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})^T \mathbf{S}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$ where $k = 1, 2, ..., K$. This term is simply zero because the innovations are independent under H_1 and $\mathbf{S}_k$ is deterministic.

Now, we derive the fifth term in (2.38). Since $\mathrm{E}(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}) = 0$, the covariance between $(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1})^T \mathbf{S}_i^{-1} (\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1})$ and $(\mathbf{z}_j - \boldsymbol{\mu}_j)^T \mathbf{R}_{u_j}^{-1} (\mathbf{z}_j - \boldsymbol{\mu}_j)$, $\forall i, j = 1, 2, ..., K$, can be derived via (2.51) as follows

$$\begin{aligned}
&\mathrm{Cov}[(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1})^T \mathbf{S}_i^{-1} (\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}), \\
&\quad (\mathbf{z}_j - \boldsymbol{\mu}_j)^T \mathbf{R}_{u_j}^{-1} (\mathbf{z}_j - \boldsymbol{\mu}_j)] \\
&= 2 \,\mathrm{tr}\left(\mathbf{S}_i^{-1} \mathbf{P}_{x\mu 1ij} \mathbf{R}_{u_j}^{-1} \mathbf{P}_{x\mu 1ij}^T \right)
\end{aligned} \tag{2.59}$$

where

$$\begin{aligned}
\mathbf{P}_{x\mu 1ij} &= \mathrm{Cov}\left(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j - \boldsymbol{\mu}_j \right) \\
&= \mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j) - \mathbf{H}_i \,\mathrm{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j \right)
\end{aligned} \tag{2.60}$$

$\mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j)$ when $i \neq j$ has been derived and provided in (2.58). $\mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j)$ when $i = j$, i.e., $\mathrm{Var}(\mathbf{z}_i)$, has been derived and provided in (2.23).

$\mathbf{H}_i \operatorname{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j\right)$ is derived as follows

$$\begin{aligned}
&\mathbf{H}_i \operatorname{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j\right) \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-1|i-1}, \mathbf{z}_j\right) \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \operatorname{Cov}\left[\left(\mathbf{I} - \mathbf{W}_{i-1}\mathbf{H}_{i-1}\right)\hat{\mathbf{x}}_{i-1|i-2} + \mathbf{W}_{i-1}\mathbf{z}_{i-1}, \mathbf{z}_j\right] \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \left(\mathbf{I} - \mathbf{W}_{i-1}\mathbf{H}_{i-1}\right) \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-1|i-2}, \mathbf{z}_j\right) \\
&\quad + \mathbf{H}_i \mathbf{F}_{i-1} \mathbf{W}_{i-1} \operatorname{Cov}\left(\mathbf{z}_{i-1}, \mathbf{z}_j\right) \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \left(\mathbf{F}_{i-2} - \mathbf{W}_{i-1}\mathbf{H}_{i-1}\mathbf{F}_{i-2}\right) \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-2|i-2}, \mathbf{z}_j\right) \\
&\quad + \mathbf{H}_i \mathbf{F}_{i-1} \mathbf{W}_{i-1} \operatorname{Cov}\left(\mathbf{z}_{i-1}, \mathbf{z}_j\right)
\end{aligned} \tag{2.61}$$

Repeating this process, we have

$$\begin{aligned}
&\mathbf{H}_i \operatorname{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j\right) \\
&= \sum_{n=1}^{i-1} \mathbf{H}_i \mathbf{F}_{i-1} \left[\prod_{m=1}^{n-1} \left(\mathbf{I} - \mathbf{W}_{i-m}\mathbf{H}_{i-m}\right) \mathbf{F}_{i-m-1}\right] \\
&\qquad \cdot \mathbf{W}_{i-n} \operatorname{Cov}\left(\mathbf{z}_{i-n}, \mathbf{z}_j\right) \\
&\quad + \mathbf{H}_i \mathbf{F}_{i-1} \left[\prod_{m=1}^{i-1} \left(\mathbf{I} - \mathbf{W}_{i-m}\mathbf{H}_{i-m}\right) \mathbf{F}_{i-m-1}\right] \\
&\qquad \cdot \operatorname{Cov}\left(\hat{\mathbf{x}}_{0|0}, \mathbf{z}_j\right) \\
&= \sum_{n=1}^{i-1} \mathbf{B}_{i,n} \mathbf{W}_{i-n} \operatorname{Cov}\left(\mathbf{z}_{i-n}, \mathbf{z}_j\right) + \mathbf{B}_{i,i} \operatorname{Cov}\left(\hat{\mathbf{x}}_{0|0}, \mathbf{z}_j\right) \\
&= \sum_{n=1}^{i-1} \mathbf{B}_{i,n} \mathbf{W}_{i-n} \operatorname{Cov}\left(\mathbf{z}_{i-n}, \mathbf{z}_j\right)
\end{aligned} \tag{2.62}$$

where $\operatorname{Cov}\left(\hat{\mathbf{x}}_{0|0}, \mathbf{z}_j\right) = \mathbf{0}$ since $\hat{\mathbf{x}}_{0|0}$ is deterministic. $\mathbf{B}_{i,n}$ has been defined in (2.31).

Finally, substituting (2.52), (2.53), (2.54) and (2.59) in (2.38), we have the following proposition:

Proposition 3 *The variance of test statistic $t_I(\mathbf{z}_{1:K})$ under H_1 when using the Kalman*

filter is

$$
\begin{aligned}
&\mathrm{Var}[t_I(\mathbf{z}_{1:K})|H_1] \\
&= 2\sum_{k=1}^{K}\Big[\mathrm{tr}\left(\mathbf{R}_{u_k}^{-1}\,\mathrm{Var}(\mathbf{z}_k-\boldsymbol{\mu}_k)\mathbf{R}_{u_k}^{-1}\,\mathrm{Var}(\mathbf{z}_k-\boldsymbol{\mu}_k)\right) \\
&\qquad +2\,\mathrm{E}(\mathbf{z}_k-\boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}\,\mathrm{Var}(\mathbf{z}_k-\boldsymbol{\mu}_k)\mathbf{R}_{u_k}^{-1}\,\mathrm{E}(\mathbf{z}_k-\boldsymbol{\mu}_k)\Big] \\
&\quad + 2\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}\Big\{\mathrm{tr}\left(\mathbf{R}_{u_i}^{-1}\mathbf{P}_{\mu 1ij}\mathbf{R}_{u_j}^{-1}\mathbf{P}_{\mu 1ij}^T\right) \\
&\qquad +2\,\mathrm{E}(\mathbf{z}_i-\boldsymbol{\mu}_i)^T\mathbf{R}_{u_i}^{-1}\mathbf{P}_{\mu 1ij}\mathbf{R}_{u_j}^{-1}\,\mathrm{E}(\mathbf{z}_j-\boldsymbol{\mu}_j)\Big\} \\
&\quad + 2Kn_z - 4\sum_{i=1}^{K}\sum_{j=1}^{K}\mathrm{tr}\left(\mathbf{S}_i^{-1}\mathbf{P}_{x\mu 1ij}\mathbf{R}_{u_j}^{-1}\mathbf{P}_{x\mu 1ij}^T\right)
\end{aligned}
\tag{2.63}
$$

2.3.3.2 Variance of Test Statistic $t_{\mathbf{I}}(\mathbf{z}_{1:K})$ Under H_0

The procedure of deriving $\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0]$ is similar to that of deriving $\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$. For simplicity, we again skip the conditioning on hypothesis H_0 in the notations in this subsection.

Under H_0, $(\mathbf{z}_k-\boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}(\mathbf{z}_k-\boldsymbol{\mu}_k)$ follows a Chi-square distribution with degree of freedom n_z. So, the first term in (2.38) is

$$
\mathrm{Var}\left[(\mathbf{z}_k-\boldsymbol{\mu}_k)^T\mathbf{R}_{u_k}^{-1}(\mathbf{z}_k-\boldsymbol{\mu}_k)\right] = 2n_z \tag{2.64}
$$

The second term in (2.38) is derived via (2.39) as follows

$$
\begin{aligned}
&\mathrm{Var}\left[\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)^T\mathbf{S}_k^{-1}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)\right] \\
&= 2\,\mathrm{tr}\left[\left(\mathbf{S}_k^{-1}\,\mathrm{Var}(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})\right)^2\right] \\
&\quad + 4\,\mathrm{E}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)^T\mathbf{S}_k^{-1}\,\mathrm{Var}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right) \\
&\qquad \cdot\,\mathbf{S}_k^{-1}\,\mathrm{E}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)
\end{aligned}
\tag{2.65}
$$

where $\mathrm{E}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)$ and $\mathrm{Var}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)$ have been derived in Section 2.3.2.2

and provided in (2.30) and (2.33), respectively.

The third term in (2.38) is the covariance between two different terms in the sequence $(\mathbf{z}_k - \boldsymbol{\mu}_k)^T \mathbf{R}_{u_k}^{-1} (\mathbf{z}_k - \boldsymbol{\mu}_k)$ where $k = 1, 2, ..., K$, which is simply zero because $\mathbf{z}_k$s are independent under H_0 and $\mathbf{R}_{u_k}$ is deterministic.

The fourth term in (2.38) can be derived via (2.51) directly as follows

$$\begin{aligned}
&\mathrm{Cov}[(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1})^T \mathbf{S}_i^{-1} (\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}), \\
&\qquad (\mathbf{z}_j - \mathbf{H}_j \hat{\mathbf{x}}_{j|j-1})^T \mathbf{S}_j^{-1} (\mathbf{z}_j - \mathbf{H}_j \hat{\mathbf{x}}_{j|j-1})] \\
&= 2\,\mathrm{tr}(\mathbf{S}_i^{-1} \mathbf{P}_{Hx0} \mathbf{S}_j^{-1} \mathbf{P}_{Hx0}^T) \\
&\quad + 4\,\mathrm{E}\left(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}^T\right)^T \mathbf{S}_i^{-1} \mathbf{P}_{Hx0} \\
&\qquad \cdot \mathbf{S}_j^{-1}\,\mathrm{E}\left(\mathbf{z}_j - \mathbf{H}_j \hat{\mathbf{x}}_{j|j-1}^T\right)
\end{aligned} \tag{2.66}$$

Since $\mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{0}$ under H_0 when $i \neq j$, $\mathbf{P}_{Hx0}$ in (2.66) can be simplified by using (2.58) and (2.62) when $i > j$ as follows

$$\begin{aligned}
\mathbf{P}_{Hx0} &= \mathrm{Cov}\left(\mathbf{z}_i - \mathbf{H}_i \hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j - \mathbf{H}_j \hat{\mathbf{x}}_{j|j-1}\right) \\
&= \mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j) + \mathbf{H}_i\,\mathrm{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \hat{\mathbf{x}}_{j|j-1}\right) \mathbf{H}_j^T \\
&\quad - \mathbf{H}_i\,\mathrm{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j\right) - \mathrm{Cov}\left(\mathbf{z}_i, \hat{\mathbf{x}}_{j|j-1}\right) \mathbf{H}_j^T \\
&= \mathbf{H}_i\,\mathrm{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \hat{\mathbf{x}}_{j|j-1}\right) \mathbf{H}_j^T - \mathbf{B}_{i,i-j} \mathbf{W}_j \mathbf{R}_{u_j}
\end{aligned} \tag{2.67}$$

When $i < j$, $\mathbf{P}_{Hx0}$ is

$$\mathbf{P}_{Hx0} = \mathbf{H}_i\,\mathrm{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \hat{\mathbf{x}}_{j|j-1}\right) \mathbf{H}_j^T - \mathbf{R}_{u_i} \mathbf{W}_i^T \mathbf{B}_{j,j-i}^T \tag{2.68}$$

where $\mathbf{H}_i \operatorname{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \hat{\mathbf{x}}_{j|j-1}\right) \mathbf{H}_j^T$ is derived as follows

$$
\begin{aligned}
&\mathbf{H}_i \operatorname{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \hat{\mathbf{x}}_{j|j-1}\right) \mathbf{H}_j^T \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-1|i-1}, \hat{\mathbf{x}}_{j-1|j-1}\right) \left(\mathbf{H}_j \mathbf{F}_{j-1}\right)^T \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \operatorname{Cov}\left[\left(\mathbf{I} - \mathbf{W}_{i-1}\mathbf{H}_{i-1}\right) \hat{\mathbf{x}}_{i-1|i-2} + \mathbf{W}_{i-1}\mathbf{z}_{i-1},\right. \\
&\quad \left.\left(\mathbf{I} - \mathbf{W}_{j-1}\mathbf{H}_{j-1}\right) \hat{\mathbf{x}}_{j-1|j-2} + \mathbf{W}_{j-1}\mathbf{z}_{j-1}\right] \left(\mathbf{H}_j \mathbf{F}_{j-1}\right)^T \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \left(\mathbf{I} - \mathbf{W}_{i-1}\mathbf{H}_{i-1}\right) \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-1|i-2}, \hat{\mathbf{x}}_{j-1|j-2}\right) \\
&\quad \cdot \left[\mathbf{H}_j \mathbf{F}_{j-1} \left(\mathbf{I} - \mathbf{W}_{j-1}\mathbf{H}_{j-1}\right)\right]^T \\
&\quad + \mathbf{H}_i \mathbf{F}_{i-1} \left(\mathbf{I} - \mathbf{W}_{i-1}\mathbf{H}_{i-1}\right) \\
&\qquad \cdot \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-1|i-2}, \mathbf{z}_{j-1}\right) \left(\mathbf{H}_j \mathbf{F}_{j-1} \mathbf{W}_{j-1}\right)^T \\
&\quad + \mathbf{H}_i \mathbf{F}_{i-1} \mathbf{W}_{i-1} \operatorname{Cov}\left(\mathbf{z}_{i-1}, \hat{\mathbf{x}}_{j-1|j-2}\right) \\
&\qquad \cdot \left[\mathbf{H}_j \mathbf{F}_{j-1} \left(\mathbf{I} - \mathbf{W}_{j-1}\mathbf{H}_{j-1}\right)\right]^T \\
&\quad + \mathbf{H}_i \mathbf{F}_{i-1} \mathbf{W}_{i-1} \operatorname{Cov}\left(\mathbf{z}_{i-1}, \mathbf{z}_{j-1}\right) \left(\mathbf{H}_j \mathbf{F}_{j-1} \mathbf{W}_{j-1}\right)^T
\end{aligned}
\tag{2.69}
$$

Repeating this procedure $\min(i,j)-1$ times, we have

$$
\begin{aligned}
&\mathbf{H}_i \operatorname{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \hat{\mathbf{x}}_{j|j-1}\right) \mathbf{H}_j^T \\
&= \mathbf{H}_i \mathbf{F}_{i-1} \left[\prod_{m=1}^{\min(i,j)-1} \left(\mathbf{I} - \mathbf{W}_{i-m}\mathbf{H}_{i-m}\right) \mathbf{F}_{i-m-1}\right] \\
&\quad \cdot \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-\min(i,j)|i-\min(i,j)}, \hat{\mathbf{x}}_{j-\min(i,j)|j-\min(i,j)}\right) \\
&\quad \cdot \left[\mathbf{H}_j \mathbf{F}_{j-1} \prod_{m=1}^{\min(i,j)-1} \left(\mathbf{I} - \mathbf{W}_{j-m}\mathbf{H}_{j-m}\right) \mathbf{F}_{j-m-1}\right]^T \\
&+ \sum_{n=1}^{\min(i,j)-1} \mathbf{H}_i \mathbf{F}_{i-1} \left[\prod_{m=1}^{n} \left(\mathbf{I} - \mathbf{W}_{i-m}\mathbf{H}_{i-m}\right) \mathbf{F}_{i-m-1}\right] \\
&\quad \cdot \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-n-1|i-n-1}, \mathbf{z}_{j-n}\right) \\
&\quad \cdot \left\{\mathbf{H}_j \mathbf{F}_{j-1} \left[\prod_{m=1}^{n-1} \left(\mathbf{I} - \mathbf{W}_{j-m}\mathbf{H}_{j-m}\right) \mathbf{F}_{j-m-1}\right] \mathbf{W}_{j-n}\right\}^T \\
&+ \sum_{n=1}^{\min(i,j)-1} \mathbf{H}_i \mathbf{F}_{i-1} \left[\prod_{m=1}^{n-1} \left(\mathbf{I} - \mathbf{W}_{i-m}\mathbf{H}_{i-m}\right) \mathbf{F}_{i-m-1}\right] \mathbf{W}_{i-n} \qquad (2.70) \\
&\quad \cdot \operatorname{Cov}\left(\mathbf{z}_{i-n}, \hat{\mathbf{x}}_{j-n-1|j-n-1}\right) \\
&\quad \cdot \left[\mathbf{H}_j \mathbf{F}_{j-1} \prod_{m=1}^{n} \left(\mathbf{I} - \mathbf{W}_{j-m}\mathbf{H}_{j-m}\right) \mathbf{F}_{j-m-1}\right]^T \\
&+ \sum_{n=1}^{\min(i,j)-1} \mathbf{H}_i \mathbf{F}_{i-1} \left[\prod_{m=1}^{n-1} \left(\mathbf{I} - \mathbf{W}_{i-m}\mathbf{H}_{i-m}\right) \mathbf{F}_{i-m-1}\right] \mathbf{W}_{i-n} \\
&\quad \cdot \operatorname{Cov}\left(\mathbf{z}_{i-n}, \mathbf{z}_{j-n}\right) \\
&\quad \cdot \left\{\mathbf{H}_j \mathbf{F}_{j-1} \left[\prod_{m=1}^{n-1} \left(\mathbf{I} - \mathbf{W}_{j-m}\mathbf{H}_{j-m}\right) \mathbf{F}_{j-m-1}\right] \mathbf{W}_{j-n}\right\}^T \\
&= \sum_{n=1}^{\min(i,j)-1} \mathbf{B}_{i,n+1} \operatorname{Cov}\left(\hat{\mathbf{x}}_{i-n-1|i-n-1}, \mathbf{z}_{j-n}\right) \mathbf{W}_{j-n}^T \mathbf{B}_{j,n}^T \\
&+ \sum_{n=1}^{\min(i,j)-1} \mathbf{B}_{i,n} \mathbf{W}_{i-n} \operatorname{Cov}\left(\mathbf{z}_{i-n}, \hat{\mathbf{x}}_{j-n-1|j-n-1}\right) \mathbf{B}_{j,n+1}^T
\end{aligned}
$$

where $\mathrm{Cov}\left(\hat{\mathbf{x}}_{i-\min(i,j)|i-\min(i,j)}, \hat{\mathbf{x}}_{j-\min(i,j)|j-\min(i,j)}\right) = \mathbf{0}$ since $\hat{\mathbf{x}}_{0|0}$ is deterministic, $\mathrm{Cov}\left(\mathbf{z}_{i-n}, \mathbf{z}_{j-n}\right) = \mathbf{0}$ since $\mathbf{z}_{i-n}$ and $\mathbf{z}_{j-n}$ are independent under H_0 when $i \neq j$, $\mathrm{Cov}\left(\hat{\mathbf{x}}_{i-n-1|i-n-1}, \mathbf{z}_{j-n}\right)$ and $\mathrm{Cov}\left(\mathbf{z}_{i-n}, \hat{\mathbf{x}}_{j-n-1|j-n-1}\right)$ can be easily derived by using (2.62).

The last term in (2.38) is derived via (2.51) as follows

$$\begin{aligned}
&\mathrm{Cov}[(\mathbf{z}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|i-1})^T\mathbf{S}_i^{-1}(\mathbf{z}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|i-1}),\\
&\quad (\mathbf{z}_j - \boldsymbol{\mu}_j)^T\mathbf{R}_{u_j}^{-1}(\mathbf{z}_j - \boldsymbol{\mu}_j)]\\
&= 2\,\mathrm{tr}\left(\mathbf{S}_i^{-1}\mathbf{P}_{x\mu 0}\mathbf{R}_{u_j}^{-1}\mathbf{P}_{x\mu 0}^T\right)\\
&\quad + 4\,\mathrm{E}\left(\mathbf{z}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|i-1}\right)^T\mathbf{S}_i^{-1}\mathbf{P}_{x\mu 0}\mathbf{R}_{u_j}^{-1}\,\mathrm{E}(\mathbf{z}_j - \boldsymbol{\mu}_j)\\
&= 2\,\mathrm{tr}\left(\mathbf{S}_i^{-1}\mathbf{P}_{x\mu 0}\mathbf{R}_{u_j}^{-1}\mathbf{P}_{x\mu 0}^T\right)
\end{aligned} \tag{2.71}$$

where $\mathbf{P}_{x\mu 0}$ is derived by using (2.62) as follows

$$\begin{aligned}
\mathbf{P}_{x\mu 0} &= \mathrm{Cov}\left(\mathbf{z}_i - \mathbf{H}_i\hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j - \boldsymbol{\mu}_j\right)\\
&= \mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j) - \mathbf{H}_i\,\mathrm{Cov}\left(\hat{\mathbf{x}}_{i|i-1}, \mathbf{z}_j\right)\\
&= \begin{cases} -\mathbf{B}_{i,i-j}\mathbf{W}_j\mathbf{R}_{u_j} & i > j\\ \mathbf{R}_{u_j} & i = j\\ \mathbf{0} & i < j \end{cases}
\end{aligned} \tag{2.72}$$

Finally, substituting (2.64), (2.65), (2.66) and (2.71) in (2.38), we have the following proposition:

Proposition 4 *The variance of test statistic $t_I(\mathbf{z}_{1:K})$ under H_0 when using the Kalman*

filter is

$$
\begin{aligned}
&\mathrm{Var}[t_I(\mathbf{z}_{1:K})|H_0] \\
&= 2\sum_{k=1}^{K}\Big\{\mathrm{tr}\left[\left(\mathbf{S}_k^{-1}\,\mathrm{Var}(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})\right)^2\right] \\
&\quad +2\,\mathrm{E}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)^T\mathbf{S}_k^{-1}\,\mathrm{Var}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right) \\
&\quad \cdot\mathbf{S}_k^{-1}\,\mathrm{E}\left(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\right)\Big\} \\
&\quad +2\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}\Big\{\mathrm{tr}\left(\mathbf{S}_i^{-1}\mathbf{P}_{Hx0}\mathbf{S}_j^{-1}\mathbf{P}_{Hx0}^T\right) \\
&\quad +2\,\mathrm{E}\left(\mathbf{z}_i-\mathbf{H}_i\hat{\mathbf{x}}_{i|i-1}^T\right)^T\mathbf{S}_i^{-1}\mathbf{P}_{Hx0} \\
&\quad \cdot\mathbf{S}_j^{-1}\,\mathrm{E}\left(\mathbf{z}_j-\mathbf{H}_j\hat{\mathbf{x}}_{j|j-1}^T\right)\Big\} \\
&\quad +2Kn_z-4\sum_{i=1}^{K}\sum_{j=1}^{K}\mathrm{tr}\left(\mathbf{S}_i^{-1}\mathbf{P}_{x\mu0}\mathbf{R}_{u_j}^{-1}\mathbf{P}_{x\mu0}^T\right)
\end{aligned}
\tag{2.73}
$$

2.4 Simulation Results

In this section, the proposed algorithms are used to solve joint sequential target detection and tracking problems, where target tracking is one case of system state estimation. The 1-dimensional target tracking has low computational complexity and could be applied in surveillance systems like intelligent robotic systems [33, 34] and visuomotor testing systems [35, 36]. Let us assume that an object is moving in a 1-dimensional space with its state at time k denoted by $\mathbf{x}_k = [\xi_k \;\; \dot{\xi}_k]^T$, where ξ_k and $\dot{\xi}_k$ are the object's position and velocity at time k, respectively. The state transition matrix is

$$\mathbf{F}_k = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}$$

for all k, where $T_s = 0.5$ seconds is the time interval between two measurements. Let us assume that there is one sensor measuring the object's position over time.

Therefore, the measurement matrix is $\mathbf{H}_k = [1 \quad 0]$ for all k. The variance matrix of state process noise is

$$\mathbf{Q}_k = 0.01 \times \begin{bmatrix} \frac{T_s^4}{4} & \frac{T_s^3}{2} \\ \frac{T_s^3}{2} & T_s^2 \end{bmatrix}$$

for all k. The mean of the object's initial state is $\hat{\mathbf{x}}_{0|0} = [0 \quad 1.5]^T$. The mean of measurement noise under H_0 is the same as the position mean of $\mathbf{x}_0$, namely $\mu = 0$. The variance of measurement noise under H_0, R_{u_k}, is the same as the position variance of $\mathbf{x}_0$, which is $\mathbf{P}_{0|0}(1,1)$ for all k. In this way, the prior estimator without measurement for the target under H_1 is not very informative. All the simulation results are based on 10^5 Monte Carlo simulations.

Firstly, we demonstrate that the theoretical results derived in Subsection 2.3.2 are correct by comparing them with the simulation results. In this simulation, the covariance matrix of sensor measurement noise is $R_{w_k} = 100$ for all k, the covariance matrix of $\mathbf{x}_0$ is $\mathbf{P}_{0|0} = diag([1000, 1])$. Both nominal probabilities of false alarm and miss are set to 10^{-3}. Threshold A is set by using its upper bound in (1.4), and threshold B is set by using its lower bound in (1.5). The results are shown in Figs. 2 and 3. It is clear that the theoretical value matches the simulation value very well under both hypotheses. From Figs. 2 and 3, we know that $\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$ is a monotonically increasing function of K, $\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0]$ is a monotonically decreasing function of K, and they cross the thresholds after 8 and 3 samples, respectively. According to (2.36) and (2.37), both $D(p(\mathbf{z}_{1:K}|H_1)||p(\mathbf{z}_{1:K}|H_0))$ and $D(p(\mathbf{z}_{1:K}|H_0)||p(\mathbf{z}_{1:K}|H_1))$ are monotonically increasing function of K. This implies, in this example, the sequential detector will eventually terminate with a very high probability.

Secondly, we demonstrate that the theoretical results derived in Subsection 2.3.3 are correct by comparing them with the simulation results. In this simulation, the settings are the same as that in the first simulation. The results are shown in Figs.

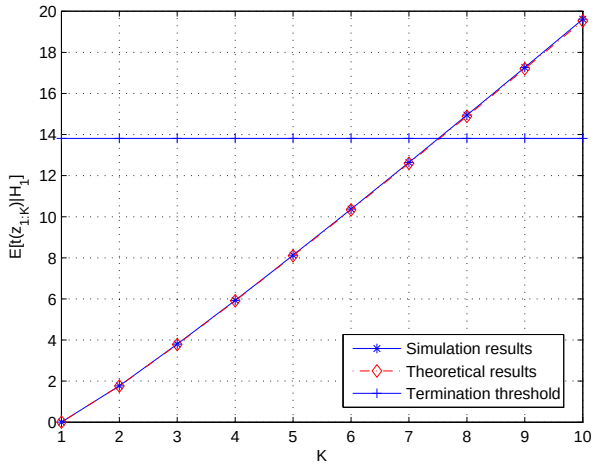

Fig. 2. $\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$ vs. K

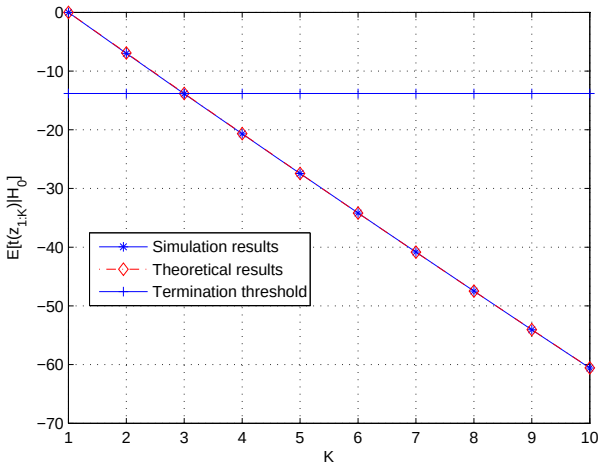

Fig. 3. $\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0]$ vs. K

4 and 5. It is clear that the theoretical value matches the simulation value very well under both hypotheses. Both $\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$ and $\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0]$ are monotonically increasing functions of K.

Thirdly, the deflection coefficient is applied to evaluate the detection performance

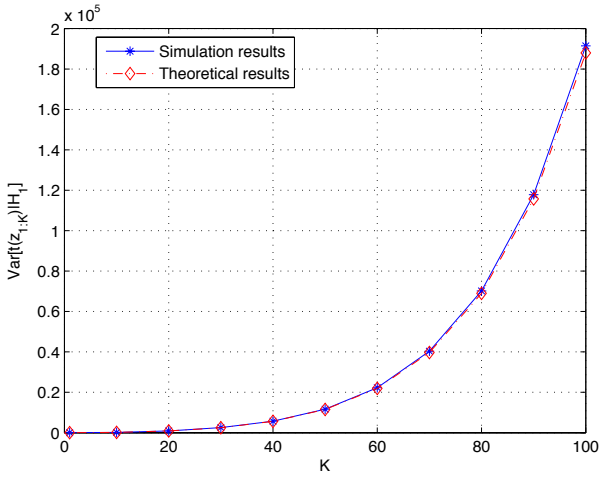

Fig. 4. $\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1]$ vs. K

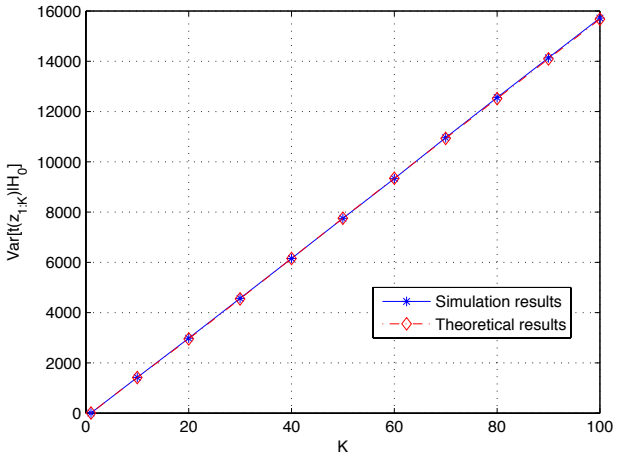

Fig. 5. $\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0]$ vs. K

of the proposed algorithm, which is defined as follows [9]

$$d^2 = \frac{(\mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_1] - \mathrm{E}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0])^2}{\mathrm{Var}[t_{\mathrm{I}}(\mathbf{z}_{1:K})|H_0]} \tag{2.74}$$

Since the distributions of $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ are not Gaussian under both hypotheses, the deflection coefficient only serves as a tractable measure of separability of the distributions of $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ under H_1 and H_0. The higher deflection coefficient is, the larger distance

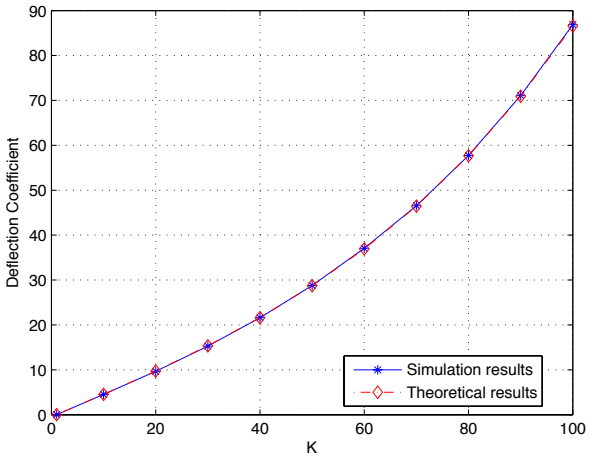

Fig. 6. Deflection Coefficient vs. K

between distributions of $t_1(\mathbf{z}_{1:K})$ under H_1 and H_0 is. The theoretical and simulation values of deflection coefficient are shown in Fig. 6. All the settings are the same as previous simulations. From Fig. 6, we know that the theoretical value matches the simulation value very well and deflection coefficient is a monotonically increasing function of K. According to the property of deflection coefficient, the distance between distributions of $t_1(\mathbf{z}_{1:K})$ under two hypotheses increases as time goes on, which implies that the probability of termination is increasing with time. The increasing distance between distributions of $t_1(\mathbf{z}_{1:K})$ under H_1 and H_0 also implies the detection performance of the propose algorithm is increasing with time.

We then evaluate the ASN required by the proposed algorithm under different SNRs in the fourth simulation. In this simulation, the SNR is defined as Fisher information about the object's position contained in $\mathbf{z}_k$. Therefore, the SNR in decibels is equal to $10\log_{10}(1/R_{w_k})$ in this case. The covariance matrix of $\hat{\mathbf{x}}_{0|0}$ is still $\mathbf{P}_{0|0} = diag([1000, 1])$. Both nominal probabilities of false alarm and miss are set to 10^{-3}. The threshold A is set by using its upper bound in (1.4), and threshold B is

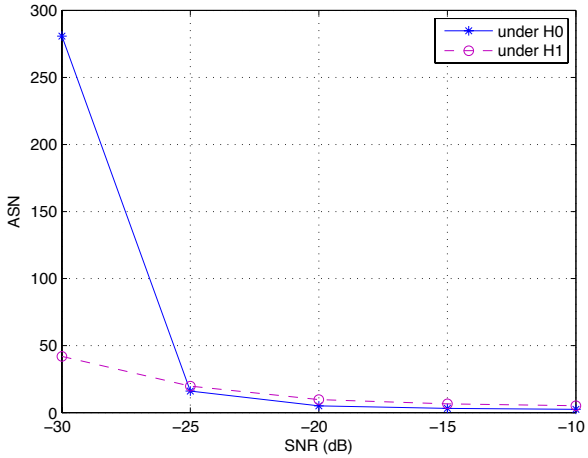

Fig. 7. ASN vs. SNR

set by using its lower bound in (1.5). As expected, from Fig. 7, we can see that the ASN decreases when SNR increases. It is equivalent to say that the ASN decreases when $\frac{R_{u_k}}{R_{w_k}}$ increases as R_{u_k} is fixed and R_{w_k} is inversely proportional to SNR. The impact of SNR on ASN under H_1 is smaller than that under H_0. As we can see, as long as SNR $\geqslant -15$dB, the proposed algorithm will terminate and make a decision very quickly, requiring a small number of samples.

Next, we evaluate the ASN with different $\mathbf{P}_{0|0}$ under both hypotheses in the fifth simulation. In this simulation, R_{w_k} is fixed at 100, the initial covariance matrix of state at time 0 is scaled by a factor κ, namely $\mathbf{P}_{0|0} = \kappa\ diag([100, 0.1])$, so $R_{u_k} = 100\kappa$. Both nominal probability of false alarm and miss are set to 10^{-3}. The threshold A is set by using its upper bound in (1.4), and threshold B is set by using its lower bound in (1.5). From Fig. 8, it is clear that ASN decreases as $\mathbf{P}_{0|0}$ increases. From this result, we also know that ASN decreases when $\frac{R_{u_k}}{R_{w_k}}$ increases, which is in agreement with the conclusion in last simulation. This can be explained as follows. Since $\mathbf{P}_{0|0}$ is already large and we already know so little about $\mathbf{x}_0$, increasing $\mathbf{P}_{0|0}$ does not make our prior knowledge about the target state at time 0 much poorer. On the other

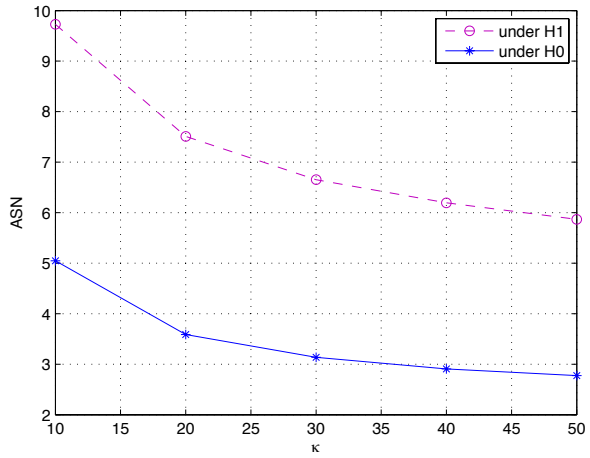

Fig. 8. ASN vs. κ

hand, under H_0 we have assumed that $R_{u_k} = \mathbf{P}_{0|0}(1,1)$. Increasing the measurement variance under H_0 (R_{u_k}) while at the same time keeping the sensor noise variance under H_1 (R_{w_k}) a constant will make it easier to discriminate one hypothesis from another, especially after the sequential detector integrates a few samples.

We evaluate the probabilities of false alarm and miss with different SNRs in the sixth simulation. The settings are the same as those in the second simulation. The comparison between the given nominal probability of false alarm α^n and the actual probability of false alarm α^a is shown in Fig. 9. It is clear that the relationship between α^n and α^a satisfies (1.6). Similarly, the relationship between β^n and β^a satisfies (1.7) as demonstrated in Fig. 10. Note that the proposed algorithm has the probabilities of error that are smaller than the nominal ones, since it will eventually terminate in this example, and the inequalities in (1.6) and (1.7) hold. From this simulation, we know that the proposed algorithm can detect an object with very low probabilities of false alarm and miss even under low SNR conditions. Analyzing the second and this simulations together, we know that the proposed algorithm can detect

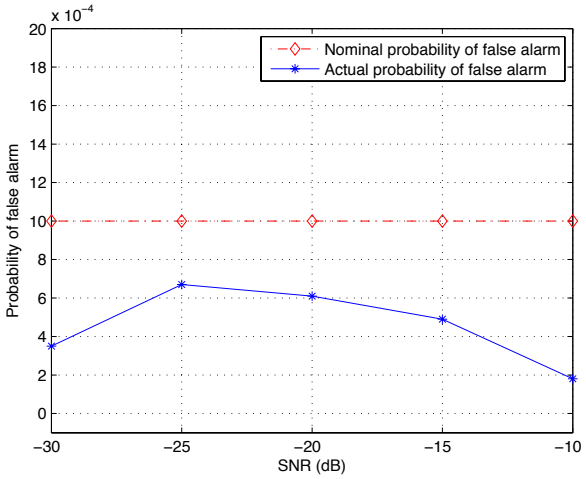

Fig. 9. Probability of False Alarm vs. SNR

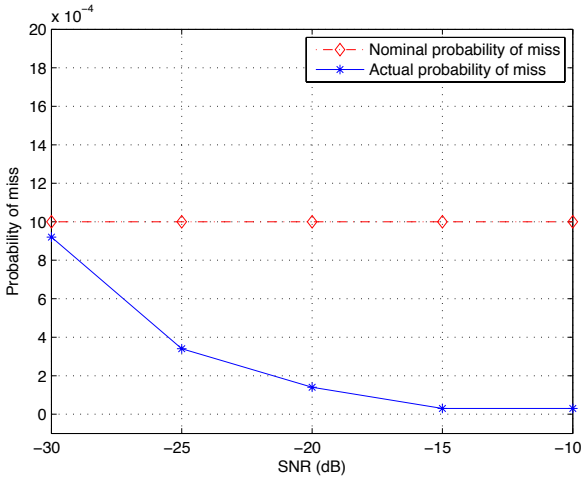

Fig. 10. Probability of Miss vs. SNR

the object with a few ASN and with very small probabilities of error under low SNR conditions.

We compare the proposed SPRT detector with the optimal FSS detector in the seventh simulation, the latter of which compares the test statistic $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ with a threshold at specified time K which is equal to the fixed sample size. The settings are the same as those in the first simulation. The detection performance of the FSS

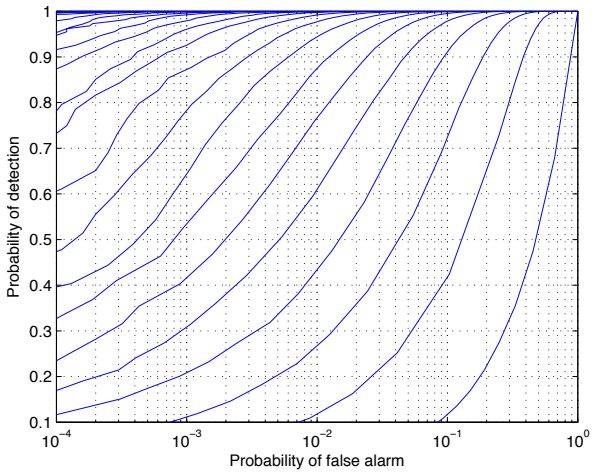

Fig. 11. ROC Curves for FSS detector

detector is demonstrated by the receiver operating characteristic (ROC) curves in Fig. 11. The ROC curves from bottom to top correspond to $K = 1,\ 2, \cdots,\ 20$, respectively. We fix the probabilities of false alarm and miss and compare the sample size required by the FSS detector with the ASN required by the SPRT detector. From the fourth simulation, we know that probability of false alarm is 6.1×10^{-4} and probability of miss is 1.4×10^{-4} when SNR is -20dB. From Fig. 11, we find that the sample size needed by the FSS detector is $K = 19$ if the FSS detector has the same or better detection performance than the SPRT detector. In the second simulation, we know that the ASN required by the proposed sequential detector under H_1 is about 10 and ASN is about 5 under H_0. Clearly, on the average, the proposed sequential detector requires a smaller number of samples than the FSS detector.

Finally, we evaluate the performance of proposed algorithm in target tracking. Under hypothesis H_1, the root mean square error (RMSE) in position or velocity over time is shown in Fig. 12. Obviously, both RMSE in position and RMSE in velocity decrease over time, which means that the accuracy of the proposed algorithm in target

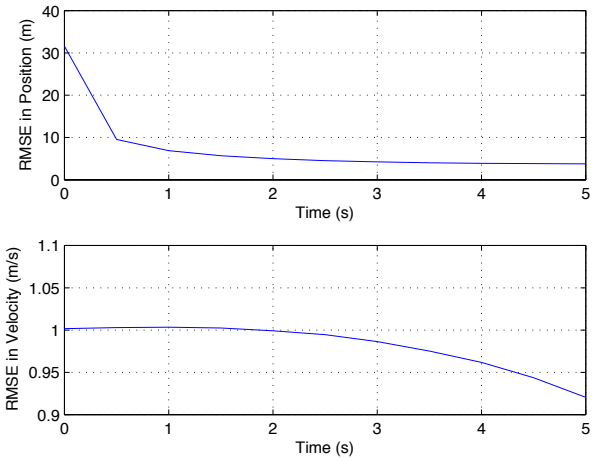

Fig. 12. RMSE in Position or Velocity over Time

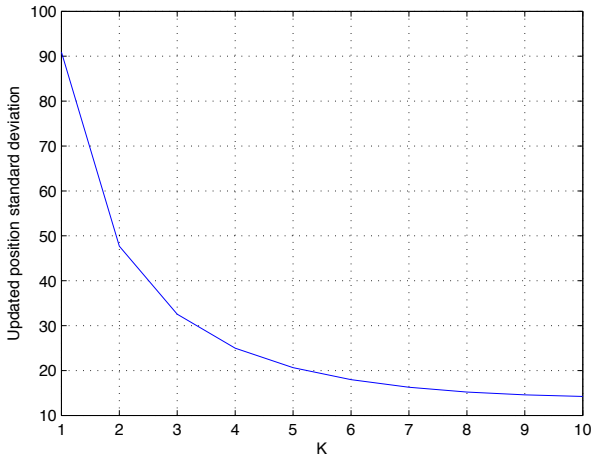

Fig. 13. Updated Position Standard Deviation vs. K

tracking is increasing over time. Under hypothesis H_1, updated position standard deviation $\sqrt{\mathbf{P}_{k|k}(1,1)}$ over time step K is shown in Fig. 13. From this figure, we know that the updated position standard deviation converges to a steady-state value within a few time steps, which means that the proposed algorithm converges to steady-state quickly.

2.5 Conclusion

In this chapter, we proposed a new joint sequential object detection and system state estimation algorithm based on the Wald's SPRT and the Kalman filter. To gain insights on the test statistic and investigate the termination of the proposed algorithm, we derived the closed-form formulas for the first and second moments of the test statistic under both hypotheses H_1 and H_0. Numerical results showed that the expected values of the test statistic are monotone functions of the number of samples, and they cross their respective thresholds in a few samples. According to the relationship between the Kullback-Leibler distance and expected value of the test statistic, the Kullback-Leibler distances are monotonically increasing functions of the number of samples which means that the distance between distributions of test statistic under H_1 and H_0 increases over time. This conclusion was also drawn by investigating the deflection coefficient, which was obtained via the expected values and variances of the test statistic. Numerical results showed that the deflection coefficient is monotonically increasing function of the number of samples which implies that both probability of termination and performance of the proposed algorithm are increasing with time. Numerical results also showed that the sequential detection algorithm detects a moving object with a small ASN and low probabilities of error even under low SNR conditions, and it outperforms the optimal FSS detector significantly in terms of the number of samples required to achieve the same detection performance. As a system state estimator, the proposed algorithm converges to steady-state quickly.

CHAPTER 3

TERMINATIVE JOINT SEQUENTIAL OBJECT DETECTION AND SYSTEM STATE ESTIMATION

3.1 Introduction

A joint sequential object detection and tracking approach based on Wald's SPRT and the Kalman filter was proposed in Chapter 2. According to [15], this approach is weakly admissible even if measurements are dependent under H_1. Moreover, it can be proved that the thresholds A, B and probabilities of error α, β still satisfy (1.4) and (1.5). However, the number of samples required by Wald's SPRT could approach to infinity, which means that these tests will not terminate with probability one. This may cause long time delay in making a decision.

In [16] and [37], Wald's SPRT was also applied when the measurements are statistically dependent. In [16], the measurements under both hypotheses were transformed into independent samples via multiplying a common transform matrix. For example, if the measurement under H_1 is a given signal with additive Gaussian noise and the measurement under H_0 is the Gaussian noise only, then the covariance matrix of measurement under H_1 and H_0 are exactly the same. The common transform matrix could be the inverse of the lower triangular matrix obtained by Cholesky decomposition of the covariance matrix. In [37], four under-sampling methods were proposed to get rid of the correlation between samples. Since the signal is given and the correlation between additive noise at different time decays as the time interval increases, the correlation between new samples will approach to zero as long as the sampling rate is low enough with respect to correlation decaying rate. Note that the

correlation decaying rate depends on SNR. The processing time will be long when SNR is low. Moreover, the signal we consider is a random process instead of a deterministic signal. Therefore, the above methods cannot be applied. The other way to make sure Wald's SPRT will terminate with probability one is truncating the procedure at $K = K_0$. However, the probabilities of error depend on the value of K_0 and the thresholds cannot be set by probabilities of error as in the case of the original SPRT.

To make sure that the sequential test procedure will terminate with probability one and the probabilities of error will be controlled by thresholds A and B, a joint terminative sequential object detection and system state estimation algorithm is proposed in this chapter. Since it was proved in [13] that Wald's SPRT procedure will terminate with probability one if the samples are independent, the proposed algorithm is designed by constructing a sequence of independent samples based on the sensor measurements. In this new algorithm, two hypothesis testing statistics are fused to guarantee that the sequential test will not only eventually terminate but also keep the power of our previous algorithm proposed in Chapter 2. The relationship between nominal probabilities of error and actual probabilities of error of fused algorithm is derived in this chapter.

This chapter is organized as follows. In Section 3.2, the terminative joint sequential object detection and tracking algorithm is proposed. In Section 3.3, the performance of the proposed algorithm is shown by simulations. Finally, this chapter is concluded in Section 3.4.

3.2 Terminative Joint Sequential Detection and System State Estimation

The joint sequential detection and system state estimation algorithm proposed in Chapter 2 is summarized as follows:

$$t_{\mathrm{I}}(\mathbf{z}_{1:K}) \begin{cases} \geqslant 2\log A & \text{stop and accept } H_1 \\ \leqslant 2\log B & \text{stop and accept } H_0 \\ \text{otherwise} & \text{continue} \end{cases} \tag{3.1}$$

where A and B are two positive constants and $B < A$, which are determined by pre-specified probabilities of false alarm and miss.

3.2.1 Independent Samples

Wald's SPRT terminates with probability one on the premise of the samples are independent under both hypotheses [13]. However, the measurements $\mathbf{z}_k$ do not satisfy this termination condition. To guarantee that Wald's SPRT will eventually terminate with probability one, we construct a sequence of independent samples $\{\mathbf{y}_l\}$ based on the measurements $\{\mathbf{z}_k\}$. To construct independent samples $\{\mathbf{y}_l\}$, we only need to consider the measurements under hypothesis H_1. Because the measurements under hypothesis H_0 are independent and identically distributed, and the linear combinations of them are still independent and identically distributed. Under hypothesis H_1, we know that the process noise sequence $\{\mathbf{v}_k\}$ and measurement noise sequence $\{\mathbf{w}_k\}$ are independent of each other. So, the measurements $\mathbf{z}_i$ and $\mathbf{z}_j$ are correlated only because they depend on the correlated states $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively. Since the measurements are linear functions of states, the independent samples can be constructed by linear combination of $\mathbf{z}_k$s in which the coefficient corresponding to $\mathbf{x}_k$ should be zero. In this chapter, the independent samples $\{\mathbf{y}_l\}$ are obtained by using three different methods.

3.2.1.1 Invertible Measurement Matrices

If $\mathbf{H}_k$s are invertible, measurement $\mathbf{z}_{k+1}$ under H_1 can be expanded as follows

$$\begin{aligned}
&\mathbf{z}_k = \mathbf{H}_k(\mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{v}_{k-1}) + \mathbf{w}_k \\
=&\mathbf{H}_k\mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{H}_k\mathbf{v}_{k-1} + \mathbf{w}_k \\
=&\mathbf{H}_k\mathbf{F}_{k-1}(\mathbf{H}_{k-1}^{-1}\mathbf{z}_{k-1} - \mathbf{H}_{k-1}^{-1}\mathbf{w}_{k-1}) + \mathbf{H}_k\mathbf{v}_{k-1} + \mathbf{w}_k \\
=&\mathbf{H}_k\mathbf{F}_{k-1}\mathbf{H}_{k-1}^{-1}\mathbf{z}_{k-1} - \mathbf{H}_k\mathbf{F}_{k-1}\mathbf{H}_{k-1}^{-1}\mathbf{w}_{k-1} + \mathbf{H}_k\mathbf{v}_{k-1} + \mathbf{w}_k
\end{aligned} \tag{3.2}$$

The independent samples can be constructed by combining two consecutive measurements as follows

$$\mathbf{y}_l^a = \mathbf{z}_{2l} - \mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1}\mathbf{z}_{2l-1} \tag{3.3}$$

According to (3.2) and (3.3), $\mathbf{y}_l^a$ under H_1 is the following function of $\mathbf{w}_{2l}$, $\mathbf{w}_{2l-1}$ and $\mathbf{v}_{2l-1}$.

$$\mathbf{y}_l^a|H_1 = \mathbf{w}_{2l} - \mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1}\mathbf{w}_{2l-1} + \mathbf{H}_{2l}\mathbf{v}_{2l-1} \tag{3.4}$$

Since $\{\mathbf{w}_k\}$ and $\{\mathbf{v}_k\}$ are white Gaussian sequences, $\mathbf{w}_k$s and $\mathbf{v}_k$s are independent for all k, and there is no common term between $\mathbf{y}_l^a$ and $\mathbf{y}_{l+1}^a$ for all $l = 1, 2, \ldots$, the sample sequence $\{\mathbf{y}_l^a\}$ under H_1 is also a white Gaussian sequence and follows $\mathcal{N}(\mathbf{0},\ \mathbf{P}_l^a|H_1)$ where $\mathbf{P}_l^a|H_1 = \mathbf{R}_{w_{2l}} + \mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1}\mathbf{R}_{w_{2l-1}}(\mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1})^T + \mathbf{H}_{2l}\mathbf{Q}_{2l-1}\mathbf{H}_{2l}^T$.

Substituting (2.3) in (3.3), $\mathbf{y}_l^a$ under H_0 is the following function of $\mathbf{u}_{2l}$ and $\mathbf{u}_{2l-1}$.

$$\mathbf{y}_l^a|H_0 = \mathbf{u}_{2l} - \mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1}\mathbf{u}_{2l-1} \tag{3.5}$$

Since $\{\mathbf{u}_k\}$ is a white Gaussian sequence and there is no common term between $\mathbf{y}_l^a|H_0$ and $\mathbf{y}_{l+1}^a|H_0$ for all $l = 1, 2, \ldots$, the sample sequence $\{\mathbf{y}_l^a\}|H_0$ is also white and follows Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_l^a,\ \mathbf{P}_l^a|H_0)$, where $\boldsymbol{\mu}_l^a = \boldsymbol{\mu}_{2l} - \mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1}\boldsymbol{\mu}_{2l-1}$ and $\mathbf{P}_l^a|H_0 = \mathbf{R}_{u_{2l}} + \mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1}\mathbf{R}_{u_{2l-1}}(\mathbf{H}_{2l}\mathbf{F}_{2l-1}\mathbf{H}_{2l-1}^{-1})^T$.

3.2.1.2 Observable Systems Based on Two Consecutive Measurements

Under H_1, we know that $\mathbf{z}_{4k-3} = \mathbf{H}_{4k-3}\mathbf{x}_{4k-3}+\mathbf{w}_{4k-3}$ and $\mathbf{z}_{4k-2} = \mathbf{H}_{4k-2}\mathbf{F}_{4k-3}\mathbf{x}_{4k-3}+ \mathbf{H}_{4k-2}\mathbf{v}_{4k-3} + \mathbf{w}_{4k-2}$. Combine these two formulas as follows

$$\begin{bmatrix} \mathbf{z}_{4k-3} \\ \mathbf{z}_{4k-2} \end{bmatrix} = \mathbf{G}_{4k-3}\mathbf{x}_{4k-3} + \begin{bmatrix} \mathbf{w}_{4k-3} \\ \mathbf{H}_{4k-2}\mathbf{v}_{4k-3} + \mathbf{w}_{4k-2} \end{bmatrix} \tag{3.6}$$

where $\mathbf{G}_{4k-3} = \begin{bmatrix} \mathbf{H}_{4k-3} \\ \mathbf{H}_{4k-2}\mathbf{F}_{4k-3} \end{bmatrix}$.

When the system is observable based on two consecutive measurements, $\mathbf{G}_{4k-3}$ is invertible and $\mathbf{x}_{4k-3}$ can be expressed as

$$\mathbf{x}_{4k-3} = \mathbf{G}_{4k-3}^{-1} \left\{ \begin{bmatrix} \mathbf{z}_{4k-3} \\ \mathbf{z}_{4k-2} \end{bmatrix} - \begin{bmatrix} \mathbf{w}_{4k-3} \\ \mathbf{H}_{4k-2}\mathbf{v}_{4k-3} + \mathbf{w}_{4k-2} \end{bmatrix} \right\} \tag{3.7}$$

It is easy to show that $\begin{bmatrix} \mathbf{z}_{4k-1} \\ \mathbf{z}_{4k} \end{bmatrix}$ can be expanded as the following function of $\mathbf{x}_{4k-3}$, $\{\mathbf{v}_k\}$, and $\{\mathbf{w}_k\}$.

$$\begin{aligned} &\begin{bmatrix} \mathbf{z}_{4k-1} \\ \mathbf{z}_{4k} \end{bmatrix} \\ &= \mathbf{G}_{4k-1}\mathbf{x}_{4k-1} + \begin{bmatrix} \mathbf{w}_{4k-1} \\ \mathbf{H}_{4k}\mathbf{v}_{4k-1} + \mathbf{w}_{4k} \end{bmatrix} \\ &= \mathbf{G}_{4k-1}\mathbf{F}_{4k-2}\mathbf{F}_{4k-3}\mathbf{x}_{4k-3} + \mathbf{G}_{4k-1}\mathbf{F}_{4k-2}\mathbf{v}_{4k-3} \\ &\quad + \mathbf{G}_{4k-1}\mathbf{v}_{4k-2} + \begin{bmatrix} \mathbf{w}_{4k-1} \\ \mathbf{H}_{4k}\mathbf{v}_{4k-1} + \mathbf{w}_{4k} \end{bmatrix} \end{aligned} \tag{3.8}$$

Substituting (3.7) in (3.8), we have

$$\begin{aligned}
&\begin{bmatrix} \mathbf{z}_{4k-1} \\ \mathbf{z}_{4k} \end{bmatrix} \\
&= \mathbf{J}_{4k-3} \left\{ \begin{bmatrix} \mathbf{z}_{4k-3} \\ \mathbf{z}_{4k-2} \end{bmatrix} - \begin{bmatrix} \mathbf{w}_{4k-3} \\ \mathbf{H}_{4k-2}\mathbf{v}_{4k-3} + \mathbf{w}_{4k-2} \end{bmatrix} \right\} \\
&\quad + \mathbf{G}_{4k-1}\mathbf{F}_{4k-2}\mathbf{v}_{4k-3} + \mathbf{G}_{4k-1}\mathbf{v}_{4k-2} \\
&\quad + \begin{bmatrix} \mathbf{w}_{4k-1} \\ \mathbf{H}_{4k}\mathbf{v}_{4k-1} + \mathbf{w}_{4k} \end{bmatrix}
\end{aligned} \tag{3.9}$$

where $\mathbf{J}_{4k-3} = \mathbf{G}_{4k-1}\mathbf{F}_{4k-2}\mathbf{F}_{4k-3}\mathbf{G}_{4k-3}^{-1}$.

Sample $\mathbf{y}_l^b$ is constructed via four consecutive measurements in $\{\mathbf{x}_k\}$ as follows

$$\mathbf{y}_l^b = \begin{bmatrix} \mathbf{z}_{4l-1} \\ \mathbf{z}_{4l} \end{bmatrix} - \mathbf{J}_{4l-3} \begin{bmatrix} \mathbf{z}_{4l-3} \\ \mathbf{z}_{4l-2} \end{bmatrix} \tag{3.10}$$

The sample $\mathbf{y}_l^b$ under H_1 is only a function of $\{\mathbf{v}_k\}$ and $\{\mathbf{w}_k\}$.

$$\begin{aligned}
&\mathbf{y}_l^b|H_1 \\
&= \mathbf{G}_{4l-1}\mathbf{F}_{4l-2}\mathbf{v}_{4l-3} - \mathbf{J}_{4l-3} \begin{bmatrix} \mathbf{w}_{4l-3} \\ \mathbf{H}_{4l-2}\mathbf{v}_{4l-3} + \mathbf{w}_{4l-2} \end{bmatrix} \\
&\quad + \mathbf{G}_{4l-1}\mathbf{v}_{4l-2} + \begin{bmatrix} \mathbf{w}_{4l-1} \\ \mathbf{H}_{4l}\mathbf{v}_{4l-1} + \mathbf{w}_{4l} \end{bmatrix}
\end{aligned} \tag{3.11}$$

Obviously, the sample sequence $\{\mathbf{y}_l^b\}|H_1$ is white and follows Gaussian distribu-

tion $\mathcal{N}(\mathbf{0},\ \mathbf{P}_l^b|H_1)$, where

$$\begin{aligned}
&\mathbf{P}_l^b|H_1 \\
&= \mathbf{G}_{4l-1}\mathbf{F}_{4l-2}\mathbf{Q}_{4l-3}\left(\mathbf{G}_{4l-1}\mathbf{F}_{4l-2}\right)^T + \mathbf{G}_{4l-1}\mathbf{Q}_{4l-2}\mathbf{G}_{4l-1}^T \\
&+ \mathbf{J}_{4l-3}\begin{bmatrix} \mathbf{R}_{w_{4l-3}} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{4l-2}\mathbf{Q}_{4l-3}\mathbf{H}_{4l-2}^T + \mathbf{R}_{w_{4l-2}} \end{bmatrix}\mathbf{J}_{4l-3}^T \\
&- \mathbf{G}_{4l-1}\mathbf{F}_{4l-2}\begin{bmatrix} \mathbf{0} & \mathbf{Q}_{4l-3}\mathbf{H}_{4l-2}^T \end{bmatrix}\mathbf{J}_{4l-3}^T \\
&+ \begin{bmatrix} \mathbf{R}_{w_{4l-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{4l}\mathbf{Q}_{4l-1}\mathbf{H}_{4l}^T + \mathbf{R}_{w_{4l}} \end{bmatrix}
\end{aligned} \tag{3.12}$$

Under H_0, the sample $\mathbf{y}_l^b$ is

$$\mathbf{y}_l^b|H_0 = \begin{bmatrix} \mathbf{u}_{4l-1} \\ \mathbf{u}_{4l} \end{bmatrix} - \mathbf{J}_{4l-3}\begin{bmatrix} \mathbf{u}_{4l-3} \\ \mathbf{u}_{4l-2} \end{bmatrix} \tag{3.13}$$

which is white and follows $\mathcal{N}(\boldsymbol{\mu}_l^b,\ \mathbf{P}_l^b|H_0)$, where $\boldsymbol{\mu}_l^b = \begin{bmatrix} \boldsymbol{\mu}_{4l-1} \\ \boldsymbol{\mu}_{4l} \end{bmatrix} - \mathbf{J}_{4l-3}\begin{bmatrix} \boldsymbol{\mu}_{4l-3} \\ \boldsymbol{\mu}_{4l-2} \end{bmatrix}$ and $\mathbf{P}_l^b|H_0 = \begin{bmatrix} \mathbf{R}_{u_{4l-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{u_{4l}} \end{bmatrix} + \mathbf{J}_{4l-3}\begin{bmatrix} \mathbf{R}_{u_{4l-3}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{u_{4l-2}} \end{bmatrix}\mathbf{J}_{4l-3}^T$.

3.2.1.3 Independent Samples Based on Cayley-Hamilton Theorem

Assume state transition matrix $\mathbf{F}_k$ and measurement matrix $\mathbf{H}_k$ are not time-varying, i.e., $\mathbf{F}_k = \mathbf{F}$ and $\mathbf{H}_k = \mathbf{H}$. It is easy to show that $\mathbf{z}_{k+m}$ is a linear function of $\mathbf{F}^m\mathbf{x}_k$ for any $m = 0, 1, 2, ...$ As long as we can find a polynomial of $\mathbf{F}$ such that it is zero, we can generate the independent samples $\{\mathbf{y}_l\}$ in the form of linear combination

of $\{\mathbf{z}_k\}$. Let us take a look at the characteristic equation of $\mathbf{F}$.

$$\begin{aligned}\kappa(\lambda) &= |\lambda\mathbf{I} - \mathbf{F}| \\ &= \lambda^{n_x} + p_1\lambda^{n_x-1} + \cdots + p_{n_x-1}\lambda + p_{n_x} \\ &= 0\end{aligned} \tag{3.14}$$

where the p_is are determined by the eigenvalues of $\mathbf{F}$. Let us denote the ith eigenvalue of $\mathbf{F}$ as λ_i. The characteristic equation of $\mathbf{F}$ can be rewritten as

$$\kappa(\lambda) = \prod_{i=1}^{n_x}(\lambda - \lambda_i) = 0 \tag{3.15}$$

According to Cayley-Hamilton theorem [38], we know that the square matrix $\mathbf{F}$ satisfies its own characteristic equation. Therefore, we have

$$\begin{aligned}\kappa(\mathbf{F}) &= \mathbf{F}^{n_x} + p_1\mathbf{F}^{n_x-1} + \cdots + p_{n_x-1}\mathbf{F} + p_{n_x}\mathbf{I} \\ &= \prod_{i=1}^{n_x}(\mathbf{F} - \lambda_i\mathbf{I}) = \mathbf{0}\end{aligned} \tag{3.16}$$

According to (3.16), the independent samples $\{\mathbf{y}_l^c\}$ is constructed as a linear combination of $n_x + 1$ adjacent measurements.

$$\begin{aligned}\mathbf{y}_l^c =& \mathbf{z}_{(n_x+1)l} + p_1\mathbf{z}_{(n_x+1)l-1} + \cdots \\ &+ p_{n_x-1}\mathbf{z}_{(n_x+1)l-(n_x-1)} + p_{n_x}\mathbf{z}_{(n_x+1)l-n_x}\end{aligned} \tag{3.17}$$

This is the general way to construct $\mathbf{y}_l^c$ for any n_x. Under H_0, $\{\mathbf{y}_l^c\}$ is a white sequence and follows Gaussian distribution as $\{\mathbf{u}_k\}$ is white and Gaussian distributed. Under H_1, the terms containing $\mathbf{x}_k$s will be canceled out by linearly combining (n_x+1)

$\mathbf{z}_k$s in this way. Substituting (2.1), (2.2), and (3.16) in (3.17), we have

$$\begin{aligned}
\mathbf{y}_l^c &= \mathbf{H}(\mathbf{F}^{n_x} + p_1\mathbf{F}^{n_x-1} + \cdots + p_{n_x}\mathbf{I})\mathbf{x}_{(n_x+1)l-n_x} \\
&\quad + g(\mathbf{w}_{(n_x+1)l}, \mathbf{w}_{(n_x+1)l-1}, \ldots, \mathbf{w}_{(n_x+1)l-n_x}) \\
&\quad + f(\mathbf{v}_{(n_x+1)l-1}, \mathbf{v}_{(n_x+1)l-2}, \ldots, \mathbf{v}_{(n_x+1)l-n_x}) \\
&= \mathbf{H}\kappa(\mathbf{F})\mathbf{x}_{(n_x+1)l-n_x} \\
&\quad + g(\mathbf{w}_{(n_x+1)l}, \mathbf{w}_{(n_x+1)l-1}, \ldots, \mathbf{w}_{(n_x+1)l-n_x}) \\
&\quad + f(\mathbf{v}_{(n_x+1)l-1}, \mathbf{v}_{(n_x+1)l-2}, \ldots, \mathbf{v}_{(n_x+1)l-n_x}) \\
&= g(\mathbf{w}_{(n_x+1)l}, \mathbf{w}_{(n_x+1)l-1}, \ldots, \mathbf{w}_{(n_x+1)l-n_x}) \\
&\quad + f(\mathbf{v}_{(n_x+1)l-1}, \mathbf{v}_{(n_x+1)l-2}, \ldots, \mathbf{v}_{(n_x+1)l-n_x})
\end{aligned} \tag{3.18}$$

where $g(\cdot)$ and $f(\cdot)$ are linear functions. Obviously, there is no common terms between $\mathbf{y}_l^c$ and $\mathbf{y}_{l-1}^c$ for all l. Also, $\mathbf{w}_k$s and $\mathbf{v}_k$s are independent, $\{\mathbf{w}_k\}$ and $\{\mathbf{v}_k\}$ are white Gaussian sequences. Therefore, the sample sequence $\{\mathbf{y}_l^c\}$ under H_1 is also white Gaussian sequence.

Let us take $n_x = 2$ as an example. In this case, $\mathbf{F}$ is a 2×2 matrix, and the characteristic equation of $\mathbf{F}$ becomes

$$\begin{aligned}
\kappa(\mathbf{F}) &= \mathbf{F}^2 + p_1\mathbf{F} + p_2\mathbf{I} \\
&= (\mathbf{F} - \lambda_1\mathbf{I})(\mathbf{F} - \lambda_2\mathbf{I}) \\
&= \mathbf{F}^2 - (\lambda_1 + \lambda_2)\mathbf{F} + \lambda_1\lambda_2\mathbf{I} = \mathbf{0}
\end{aligned} \tag{3.19}$$

where $p_1 = -(\lambda_1 + \lambda_2)$ and $p_2 = \lambda_1\lambda_2$.

According to (3.19), the sample $\mathbf{y}_l^c$ in (3.17) becomes

$$\begin{aligned}
\mathbf{y}_l^c &= \mathbf{z}_{3l} + p_1\mathbf{z}_{3l-1} + p_2\mathbf{z}_{3l-2} \\
&= \mathbf{z}_{3l} - (\lambda_1 + \lambda_2)\mathbf{z}_{3l-1} + \lambda_1\lambda_2\mathbf{z}_{3l-2}
\end{aligned} \tag{3.20}$$

$\mathbf{z}_{3l}$, $\mathbf{z}_{3l-1}$, and $\mathbf{z}_{3l-2}$ are expanded as functions of $\mathbf{x}_{3l-2}$ as follows

$$\begin{aligned}\mathbf{z}_{3l} &= \mathbf{H}\mathbf{x}_{3l} + \mathbf{w}_{3l} \\ &= \mathbf{H}(\mathbf{F}^2\mathbf{x}_{3l-2} + \mathbf{F}\mathbf{v}_{3l-2} + \mathbf{v}_{3l-1}) + \mathbf{w}_{3l}\end{aligned} \tag{3.21}$$

$$\begin{aligned}\mathbf{z}_{3l-1} &= \mathbf{H}\mathbf{x}_{3l-1} + \mathbf{w}_{3l-1} \\ &= \mathbf{H}(\mathbf{F}\mathbf{x}_{3l-2} + \mathbf{v}_{3l-2}) + \mathbf{w}_{3l-1}\end{aligned} \tag{3.22}$$

and

$$\mathbf{z}_{3l-2} = \mathbf{H}\mathbf{x}_{3l-2} + \mathbf{w}_{3l-2} \tag{3.23}$$

Substituting (3.21), (3.22), and (3.23) in (3.20), we have

$$\begin{aligned}\mathbf{y}_l^c &= \mathbf{z}_{3l} - (\lambda_1 + \lambda_2)\mathbf{z}_{3l-1} + \lambda_1\lambda_2\mathbf{z}_{3l-2} \\ &= \mathbf{H}\left\{\left[\mathbf{F}^2 - (\lambda_1 + \lambda_2)\mathbf{F} + \lambda_1\lambda_2\mathbf{I}\right]\mathbf{x}_{3l-2}\right. \\ &\quad \left. + \left[\mathbf{F} - (\lambda_1 + \lambda_2)\mathbf{I}\right]\mathbf{v}_{3l-2} + \mathbf{v}_{3l-1}\right\} \\ &\quad + \mathbf{w}_{3l} - (\lambda_1 + \lambda_2)\mathbf{w}_{3l-1} + \lambda_1\lambda_2\mathbf{w}_{3l-2}\end{aligned} \tag{3.24}$$

where the coefficient of $\mathbf{x}_{3l-2}$ is equal to zero as shown in (3.19). So, the sample $\mathbf{y}_l^c$ under H_1 is as follows

$$\begin{aligned}\mathbf{y}_l^c|H_1 = \mathbf{H}&\left\{\left[\mathbf{F} - (\lambda_1 + \lambda_2)\mathbf{I}\right]\mathbf{v}_{3l-2} + \mathbf{v}_{3l-1}\right\} \\ &+ \mathbf{w}_{3l} - (\lambda_1 + \lambda_2)\mathbf{w}_{3l-1} + \lambda_1\lambda_2\mathbf{w}_{3l-2}\end{aligned} \tag{3.25}$$

Now, we derive the distribution of independent samples $\mathbf{y}_l^c$ under H_1 and H_0, respectively. Let $\mathbf{P}_l^c|H_1$ and $\mathbf{P}_l^c|H_0$ denote the covariance matrices under H_1 and H_0, respectively. According to (3.25), we get $\mathbf{P}_l^c|H_1$ as follows

$$\begin{aligned}\mathbf{P}_l^c|H_1 = \mathbf{H}&\left[\mathbf{F} - (\lambda_1 + \lambda_2)\mathbf{I}\right]\mathbf{Q}_{3l-2}\left[\mathbf{F} - (\lambda_1 + \lambda_2)\mathbf{I}\right]^T\mathbf{H}^T \\ &+ \mathbf{H}\mathbf{Q}_{3l-1}\mathbf{H}^T + \mathbf{R}_{w_{3l}} + (\lambda_1 + \lambda_2)^2\mathbf{R}_{w_{3l-1}} + \lambda_1^2\lambda_2^2\mathbf{R}_{w_{3l-2}}\end{aligned} \tag{3.26}$$

Obviously, the mean of $\mathbf{y}_l^c$ under H_1 is zero. Therefore, $\mathbf{y}_l^c|H_1 \sim \mathcal{N}(\mathbf{0},\ \mathbf{P}_l^c|H_1)$.

$\mathbf{y}_l^c$ under H_0 is obtained in the same way as (3.20) and we have

$$\begin{aligned}\mathbf{y}_l^c|H_0 &= \mathbf{z}_{3l} - (\lambda_1+\lambda_2)\mathbf{z}_{3l-1} + \lambda_1\lambda_2\mathbf{z}_{3l-2} \\ &= \mathbf{u}_{3l} - (\lambda_1+\lambda_2)\mathbf{u}_{3l-1} + \lambda_1\lambda_2\mathbf{u}_{3l-2}\end{aligned} \tag{3.27}$$

It's easy to show that $\mathbf{y}_l^c|H_0 \sim \mathcal{N}(\boldsymbol{\mu}_l^c,\ \mathbf{P}_l^c|H_0)$ where

$$\boldsymbol{\mu}_l^c = \boldsymbol{\mu}_{3l} - (\lambda_1+\lambda_2)\boldsymbol{\mu}_{3l-1} + \lambda_1\lambda_2\boldsymbol{\mu}_{3l-2} \tag{3.28}$$

and

$$\mathbf{P}_l^c|H_0 = \mathbf{R}_{u_{3l}} + (\lambda_1+\lambda_2)^2\mathbf{R}_{u_{3l-1}} + \lambda_1^2\lambda_2^2\mathbf{R}_{u_{3l-2}} \tag{3.29}$$

For all of the three methods, the mean and variance of $\mathbf{y}_l$ can be calculated off-line.

3.2.2 Fused Hypothesis Testing Statistic

No matter which method is chosen to construct independent sample $\mathbf{y}_l$, the distributions of $\mathbf{y}_l$ under both hypotheses are known. So, the log-likelihood ratio is adopted to generate the hypothesis testing statistic. Let $t_{\mathrm{II}}(\mathbf{y}_{1:L})$ denote the hypothesis testing statistic for the samples accumulated up to the Lth step based on the

independent samples. The test statistic t_{II} based on independent samples $\{\mathbf{y}_l\}$ is

$$\begin{aligned} & t_{\mathrm{II}}(\mathbf{y}_{1:L}) \\ & = 2\log\frac{p(\mathbf{y}_{1:L}|H_1)}{p(\mathbf{y}_{1:L}|H_0)} = 2\log\frac{\prod_{l=1}^{L} p(\mathbf{y}_l|H_1)}{\prod_{l=1}^{L} p(\mathbf{y}_l|H_0)} \\ & = 2\sum_{l=1}^{L}\log\frac{p(\mathbf{y}_l|H_1)}{p(\mathbf{y}_l|H_0)} \\ & = \sum_{l=1}^{L}\Big[\log\frac{|\mathbf{P}_l|H_0|}{|\mathbf{P}_l|H_1|} - (\mathbf{y}_l)^T(\mathbf{P}_l|H_1)^{-1}\mathbf{y}_l \\ & \qquad +(\mathbf{y}_l-\boldsymbol{\mu}_l)^T(\mathbf{P}_l|H_0)^{-1}(\mathbf{y}_l-\boldsymbol{\mu}_l)\Big] \end{aligned} \tag{3.30}$$

where $L = 1, 2, \cdots$.

To guarantee that the joint sequential detection and system state estimation algorithm in Chapter 2 will eventually terminate with probability one, a terminative algorithm is constructed as follows by applying a fused hypothesis testing statistic.

1. Hypothesis H_1 will be accepted and the sequential test will terminate if

$$U_K \geqslant 2\log A \tag{3.31}$$

where

$$U_K = \max\{t_{\mathrm{I}}(\mathbf{z}_{1:K}),\ t_{\mathrm{II}}(\mathbf{y}_{1:\lfloor K/n_s \rfloor})\} \tag{3.32}$$

2. Hypothesis H_0 will be accepted and the sequential test will terminate if

$$L_K \leqslant 2\log B \tag{3.33}$$

where

$$L_K = \min\{t_{\mathrm{I}}(\mathbf{z}_{1:K}),\ t_{\mathrm{II}}(\mathbf{y}_{1:\lfloor K/n_s \rfloor})\} \tag{3.34}$$

3. Otherwise, the sequential test will continue to take the next sample.

Note that n_s denotes the number of measurements $\mathbf{z}_k$ used to generate each indepen-

dent sample $\mathbf{y}_l$. The value of n_s depends on which method is used to generate $\mathbf{y}_l$: $n_s = 2$ when the first method in Section 3.2.1.1 is adopted, $n_s = 4$ when the second method in Section 3.2.1.2 is adopted, and $n_s = n_x + 1$ when the third method in Section 3.2.1.3 is adopted.

In this procedure, $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ is calculated for each positive integer K, and $t_{\mathrm{II}}(\mathbf{y}_{1:K/n_s})$ is fused together with $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ when $K = n_s \times r, r = 1, 2, \cdots$. This procedure will terminate if either the larger one between $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ and $t_{\mathrm{II}}(\mathbf{y}_{1:K/n_s})$ is greater than or equal to $2\log A$ or the smaller one is less than or equal to $2\log B$ when $K = n_s \times r, r = 1, 2, \cdots$. When $K \neq n_s \times r$, equivalently only $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ is used to make decision as $t_{\mathrm{II}}(\mathbf{y}_{1:\lfloor K/n_s \rfloor})$ does not provide any new information. If we only use $t_{\mathrm{II}}(\mathbf{y}_{1:\lfloor K/n_s \rfloor})$ as hypothesis testing statistic in Wald's SPRT, it will eventually terminate as $\mathbf{y}_l$s are independent. It is easy to show that the proposed algorithm will also terminate with probability one since Wald's SPRT procedure will terminate as long as either $t_{\mathrm{I}}(\mathbf{z}_{1:K})$ or $t_{\mathrm{II}}(\mathbf{y}_{1:\lfloor K/n_s \rfloor})$ crosses one threshold.

Now, for the algorithm based on fused statistics, we derive the upper bound on α_f^a and β_f^a in terms of α_f^n and β_f^n, where α_f^a and β_f^a denote the actual probabilities of false alarm and miss of the proposed fused algorithm respectively, and α_f^n and β_f^n denote the nominal probabilities of false alarm and miss of the proposed fused algorithm respectively.

We know that the thresholds A and B, actual probability of false alarm α^a, and actual probability of miss β^a will satisfy the following inequalities when the samples are independent [13].

$$A \leqslant \frac{1 - \beta^a}{\alpha^a} \tag{3.35}$$

$$B \geqslant \frac{\beta^a}{1 - \alpha^a} \tag{3.36}$$

Since the proposed fused algorithm will eventually terminate, denote the largest

number of samples required by it as N_s. Note that the proposed fused algorithm contains two sequential tests with only one test statistic within limited steps N_s: the algorithm based on dependent measurements in Chapter 2 and the algorithm based on independent samples which only uses t_{II} in Wald's SPRT. First, we show that the inequalities (3.35) and (3.36) always hold for any sequential test with only one test statistic within limited steps (no matter whether the test statistic crosses the thresholds or not). To simplify the notation, denote a collection of the samples under consideration as $\mathbf{x} := [x_1, x_2, \ldots, x_k]^T$ where $1 \leqslant k \leqslant N_s$. Denote the joint PDFs of $\mathbf{x}$ under H_1 and H_0 are $p_{1k}(\mathbf{x})$ and $p_{0k}(\mathbf{x})$, respectively. Therefore, the likelihood ratio is $\Lambda_k = p_{1k}(\mathbf{x})/p_{0k}(\mathbf{x})$. The decision set of accepting H_1 is denoted by $R_1 := \{\mathbf{x} : \Lambda_k \geqslant A, B < \Lambda_m < A, \forall m = 1, 2, \ldots, k-1\}$. The decision set of accepting H_0 as $R_0 := \{\mathbf{x} : \Lambda_k \leqslant B, B < \Lambda_m < A, \forall m = 1, 2, \ldots, k-1\}$. If $\mathbf{x}$ does not fall in either R_1 or R_0 when $k = N_s$, Wald's SPRT is forced to terminate and no decision is made. Note that this procedure is different from truncated Wald's SPRT which is forced to make a decision when $k = N_s$. Since $\mathbf{x}$ may not fall in either R_1 or R_0 when $k = N_s$, the probability of detection is less than or equal to $1 - \beta^a$. So, we have

$$\begin{aligned} 1 - \beta^a &\geqslant \int_{R_1} p_{1k}(\mathbf{x}) d\mathbf{x} = \int_{R_1} \frac{p_{1k}(\mathbf{x})}{p_{0k}(\mathbf{x})} p_{0k}(\mathbf{x}) d\mathbf{x} \\ &\geqslant A \int_{R_1} p_{0k}(\mathbf{x}) d\mathbf{x} = A\alpha^a \end{aligned} \tag{3.37}$$

that is, $1 - \beta^a \geqslant A\alpha^a$ which is equivalent to (3.35). Note that if there is no limit on the number of samples N_s, then $1 - \beta^a$ equals to probability of detection. In this case, the inequality (3.35) is still true. However, if the Wald's SPRT is truncated, inequality (3.35) is not true.

Similarly, we have

$$\begin{aligned} 1-\alpha^a &\geqslant \int_{R_0} p_{0k}(\mathbf{x})d\mathbf{x} = \int_{R_0} \frac{p_{0k}(\mathbf{x})}{p_{1k}(\mathbf{x})} p_{1k}(\mathbf{x})d\mathbf{x} \\ &\geqslant \frac{1}{B}\int_{R_0} p_{1k}(\mathbf{x})d\mathbf{x} = \frac{1}{B}\beta^a \end{aligned} \tag{3.38}$$

that is, $1-\alpha^a \geqslant \beta^a/B$ which is equivalent to (3.36), which is true no matter whether the number of samples is limited or not. However, (3.36) is not true if the Wald's SPRT is truncated.

If we set the thresholds as $A = \frac{1-\beta^n}{\alpha^n}$ and $B = \frac{\beta^n}{1-\alpha^n}$, then we obtain the following inequalities from (3.35) and (3.36).

$$\frac{\alpha^a}{1-\beta^a} \leqslant \frac{\alpha^n}{1-\beta^n} \tag{3.39}$$

and

$$\frac{\beta^a}{1-\alpha^a} \leqslant \frac{\beta^n}{1-\alpha^n} \tag{3.40}$$

where α^n and β^n denote the nominal probability of false alarm and nominal probability of miss, respectively.

Since $0 < 1-\alpha^a < 1$ and $0 < 1-\beta^a < 1$, (3.39) and (3.40) become

$$\alpha^a \leqslant \frac{\alpha^n}{1-\beta^n} \tag{3.41}$$

and

$$\beta^a \leqslant \frac{\beta^n}{1-\alpha^n} \tag{3.42}$$

which are the same as (1.6) and (1.7). Multiplying (3.39) by $(1-\beta^n)(1-\beta^a)$ and (3.40) by $(1-\alpha^n)(1-\alpha^a)$ and adding them, we obtain the following relationship between actual probabilities of error and nominal probabilities of error.

$$\alpha^a + \beta^a \leqslant \alpha^n + \beta^n \tag{3.43}$$

Up to now, we have proved that the following proposition is true no matter whether the test statistic is generated by independent samples or dependent samples.

Proposition 5 *For any SPRT with fixed thresholds A and B where $B < A$, if only one test statistic is adopted and the SPRT is not truncated, then*

1. *the thresholds A and B, the actual probability of false alarm α^a, and the actual probability of miss β^a satisfy $A \leqslant \frac{1-\beta^a}{\alpha^a}$ and $B \geqslant \frac{\beta^a}{1-\alpha^a}$,*
2. *suppose $A = \frac{1-\beta^n}{\alpha^n}$ and $B = \frac{\beta^n}{1-\alpha^n}$, the upper bounds on the actual probability of false alarm α^a and the actual probability of miss β^a are $\alpha^a \leqslant \frac{\alpha^n}{1-\beta^n}$ and $\beta^a \leqslant \frac{\beta^n}{1-\alpha^n}$,*
3. *the actual probabilities of error and nominal probabilities of error have the relationship $\alpha^a + \beta^a \leqslant \alpha^n + \beta^n$.*

Let α_{I}^a and β_{I}^a denote the actual probabilities of false alarm and miss of the dependent algorithm in Chapter 2 respectively. Let α_{II}^a and β_{II}^a denote the actual probabilities of false alarm and miss of the independent algorithm which only uses t_{II} in Wald's SPRT respectively. Let α^n and β^n be the nominal probability of false alarm and miss of the SPRT using t_{I} or t_{II} as test statistic respectively. If we still use $A = \frac{1-\beta^n}{\alpha^n}$ and $B = \frac{\beta^n}{1-\alpha^n}$ as the thresholds in the proposed fused algorithm, we will obtain the following inequalities from (3.43).

$$\alpha_{\mathrm{I}}^a + \beta_{\mathrm{I}}^a \leqslant \alpha^n + \beta^n \tag{3.44}$$

and

$$\alpha_{\mathrm{II}}^a + \beta_{\mathrm{II}}^a \leqslant \alpha^n + \beta^n \tag{3.45}$$

Adding (3.44) to (3.45), we have

$$\alpha_{\mathrm{I}}^a + \beta_{\mathrm{I}}^a + \alpha_{\mathrm{II}}^a + \beta_{\mathrm{II}}^a \leqslant 2\alpha^n + 2\beta^n \tag{3.46}$$

Now, let us find the relationship between $\alpha_{\mathrm{I}}^a + \alpha_{\mathrm{II}}^a$ and the actual probability of false alarm of the proposed fused algorithm α_f^a. Denote t_{I} and t_{II} corresponding to U_k and L_k by t_{I}^k and t_{II}^k respectively. The relationship between α_f^a and $\alpha_{\mathrm{I}}^a + \alpha_{\mathrm{II}}^a$ is derived as follows

$$\begin{aligned}
\alpha_f^a &= \sum_{k=1}^{N_s} \Pr\{U_k \geqslant 2\log A, \\
&\quad 2\log B < \{L_i\}_{i=1}^{k-1} \leqslant \{U_i\}_{i=1}^{k-1} < 2\log A | H_0\} \\
&= \sum_{k=1}^{N_s} \Pr\{\max(t_{\mathrm{I}}^k, t_{\mathrm{II}}^k) \geqslant 2\log A, \\
&\quad (2\log B < \{t_{\mathrm{I}}^i\}_{i=1}^{k-1}, \{t_{\mathrm{II}}^i\}_{i=1}^{k-1} < 2\log A) | H_0\} \\
&= \sum_{k=1}^{N_s} \Pr\{(t_{\mathrm{I}}^k \geqslant 2\log A \text{ or } t_{\mathrm{II}}^k \geqslant 2\log A), \\
&\quad (2\log B < \{t_{\mathrm{I}}^i\}_{i=1}^{k-1}, \{t_{\mathrm{II}}^i\}_{i=1}^{k-1} < 2\log A) | H_0\} \\
&\leqslant \sum_{k=1}^{N_s} \Pr\{t_{\mathrm{I}}^k \geqslant 2\log A, 2\log B < \{t_{\mathrm{I}}^i\}_{i=1}^{k-1} < 2\log A, \\
&\qquad 2\log B < \{t_{\mathrm{II}}^i\}_{i=1}^{k-1} < 2\log A | H_0\} \\
&\quad + \sum_{k=1}^{N_s} \Pr\{t_{\mathrm{II}}^k \geqslant 2\log A, 2\log B < \{t_{\mathrm{I}}^i\}_{i=1}^{k-1} < 2\log A, \\
&\qquad 2\log B < \{t_{\mathrm{II}}^i\}_{i=1}^{k-1} < 2\log A | H_0\} \\
&\leqslant \sum_{k=1}^{N_s} \Pr\{t_{\mathrm{I}}^k \geqslant 2\log A, 2\log B < \{t_{\mathrm{I}}^i\}_{i=1}^{k-1} < 2\log A | H_0\} \\
&\quad + \sum_{k=1}^{N_s} \Pr\{t_{\mathrm{II}}^k \geqslant 2\log A, 2\log B < \{t_{\mathrm{II}}^i\}_{i=1}^{k-1} < 2\log A | H_0\} \\
&= \alpha_{\mathrm{I}}^a + \alpha_{\mathrm{II}}^a
\end{aligned} \tag{3.47}$$

Similarly, the actual probability of miss of the proposed fused algorithm β_f^a has the following relationship with β_{I}^a and β_{II}^a.

$$\beta_f^a \leqslant \beta_{\mathrm{I}}^a + \beta_{\mathrm{II}}^a \tag{3.48}$$

Adding (3.47) to (3.48), we have

$$\alpha_f^a + \beta_f^a \leqslant \alpha_{\mathrm{I}}^a + \beta_{\mathrm{I}}^a + \alpha_{\mathrm{II}}^a + \beta_{\mathrm{II}}^a \tag{3.49}$$

Let α_f^n and β_f^n be the nominal probabilities of false alarm and miss of the proposed fused algorithm, respectively. Let us assume that $\alpha_f^n = 2\alpha^n$ and $\beta_f^n = 2\beta^n$. Then, $A = \frac{2-\beta_f^n}{\alpha_f^n}$ and $B = \frac{\beta_f^n}{2-\alpha_f^n}$. $\alpha_f^a + \beta_f^a$ has the following relationship with $\alpha_f^n + \beta_f^n$ by comparing (3.46) with (3.49).

$$\alpha_f^a + \beta_f^a \leqslant \alpha_f^n + \beta_f^n \tag{3.50}$$

That is, the sum of actual probabilities of false alarm and miss will be upper bounded by the sum of nominal probabilities of false alarm and miss.

Furthermore, we can find the upper bound of α_f^a via (3.47) and (3.41) as follows

$$\alpha_f^a \leqslant \alpha_{\mathrm{I}}^a + \alpha_{\mathrm{II}}^a \leqslant \frac{\alpha^n}{1-\beta^n} + \frac{\alpha^n}{1-\beta^n} = \frac{\alpha_f^n}{1-\frac{\beta_f^n}{2}} \tag{3.51}$$

Similarly, we have

$$\beta_f^a \leqslant \beta_{\mathrm{I}}^a + \beta_{\mathrm{II}}^a \leqslant \frac{\beta^n}{1-\alpha^n} + \frac{\beta^n}{1-\alpha^n} = \frac{\beta_f^n}{1-\frac{\alpha_f^n}{2}} \tag{3.52}$$

It's easy to prove that $\frac{\alpha_f^n}{1-\frac{\beta_f^n}{2}} < \alpha_f^n + \beta_f^n$ and $\frac{\beta_f^n}{1-\frac{\alpha_f^n}{2}} < \alpha_f^n + \beta_f^n$. Therefore, the feasible region of α_f^a and β_f^a is shown as in Fig. 14.

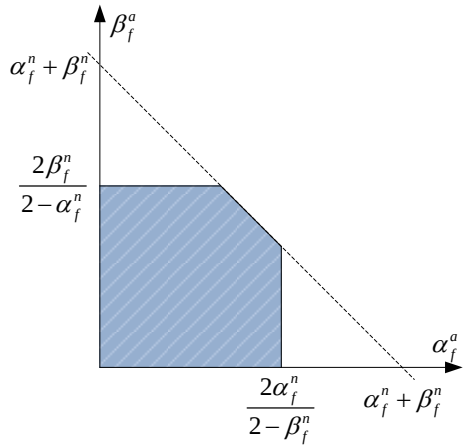

Fig. 14. The Feasible Region of α_f^a and β_f^a

3.3 Simulation Results

In this subsection, we evaluate performance of the algorithm proposed in Chapter 3. Most settings are the same as that in Chapter 2. The covariance matrix of the object's initial state is $\mathbf{P}_{0|0} = diag([1000, 1])$. The SNR is defined as Fisher information about the object's position contained in $\mathbf{z}_k$. Therefore, the SNR in decibels is equal to $10\log_{10}(1/R_{w_k})$ in this case. Let $\alpha_f^n = \beta_f^n = 10^{-3}$. The thresholds A and B are set via $A = \frac{2-\beta_f^n}{\alpha_f^n}$ and $B = \frac{\beta_f^n}{2-\alpha_f^n}$. Since $\mathbf{H} = [1 \;\; 0]$ is not invertible but the system is observable, $\mathbf{y}_l^b$ and $\mathbf{y}_l^c$ are used to generate independent samples. For simplicity, the algorithm proposed in Chapter 2 is named dependent algorithm, the Wald's SPRT that only uses independent samples $\mathbf{y}_l^c$ is named independent algorithm I, the Wald's SPRT that only uses independent samples $\mathbf{y}_l^b$ is named independent algorithm II, the Wald's SPRT uses fused samples generated by $\mathbf{y}_l^c$ is named fused algorithm I, the Wald's SPRT uses fused samples generated by $\mathbf{y}_l^b$ is named fused algorithm II. Since $n_x = 2$ in this simulation, every three consecutive measurements are used to generate independent samples $\mathbf{y}_l^c$. The eigenvalues of $\mathbf{F}$ are $\lambda_1 = \lambda_2 = 1$.

First, we evaluate the ASN required by the five algorithms under different SNRs. The simulation results under H_0 and under H_1 are shown in Table 2 and Table 3, respectively. From the results, we know that ASN is inversely proportional to SNR under both hypotheses. This is just as expected. We know that when the SNR increases, the distance between distributions of samples under two hypotheses also increases. So, less time is needed to terminate under higher SNR. From the results, we also know that the ASNs required by the fused algorithms are always less than those required by the dependent and independent algorithms. This is because the fused algorithm uses both the test statistics of dependent and independent algorithms and it terminates as long as one test statistic crosses the thresholds. The ASNs required by the independent algorithms are much higher than the other two algorithms as their test statistics contain less information when eliminating the terms containing $\mathbf{x}_k$. By designing the fused algorithm, the Wald's SPRT will terminate with probability one and the performance of dependent algorithm in terms of ASN will be slightly improved.

Table 2. ASNs of five algorithms under H_0

SNR (dB)	-25	-20	-15	-10	-5
Dependent algorithm	17.3395	5.2373	3.1512	2.5144	2.2594
Independent algorithm I	55.5248	13.0013	6.4813	4.5392	3.7619
Fused algorithm I	17.3014	5.2166	3.1419	2.5107	2.2578
Independent algorithm II	38.5708	9.8977	5.6360	4.5220	4.1728
Fused algorithm II	17.2876	5.2199	3.1465	2.5128	2.2590

Table 3. ASNs of five algorithms under H_1

SNR (dB)	-25	-20	-15	-10	-5
Dependent algorithm	21.1927	10.5060	7.0246	5.4581	4.6130
Independent algorithm I	100.3770	34.5868	20.1028	14.1998	11.0138
Fused algorithm I	21.1878	10.5054	7.0243	5.4581	4.6130
Independent algorithm II	68.4534	24.1695	14.4066	10.2750	8.3947
Fused algorithm II	21.1660	10.5007	7.0230	5.4570	4.6129

Second, we evaluate the relationship between α_f^a, β_f^a, and their upper bounds.

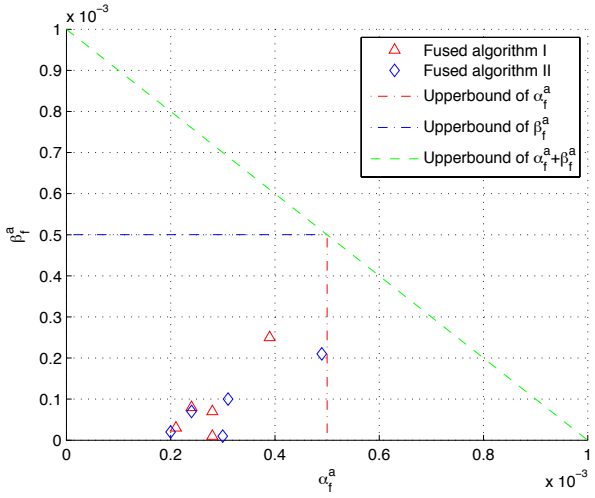

Fig. 15. Compare (α_f^a, β_f^a) Pair and Its Feasible Region under Different SNRs

The (α_f^a, β_f^a) pairs under different SNRs and the feasible region are plotted in Fig. 15. Obviously, relationships (3.50), (3.51) and (3.52) derived in Chapter 3 are satisfied. Therefore, the thresholds in the fused algorithm can be set by the nominal probabilities of error and the upper bounds on the actual probabilities of error can be obtained via nominal probabilities of error.

Third, we compare the probabilities of error made by the five algorithms. The probability of false alarm and probability of miss at different SNRs are shown in Fig. 16 and Fig. 17, respectively. For each independent algorithm, the actual probabilities of error of are less than the nominal probabilities of error $\alpha^n = \beta^n = 0.5 \times 10^{-3}$ respectively, which means that (3.41), (3.42) and (3.43) are satisfied. In Fig. 16, the summation of probabilities of false alarm of dependent algorithm and independent algorithm is greater than probability of false alarm of the fused algorithm, which satisfies (3.47). Similarly, (3.48) is demonstrated by Fig. 17. In Fig. 16, the probability of false alarm of the fused algorithm is slightly greater than that of the dependent algorithm. This is also the case for the probability of miss. This is because every

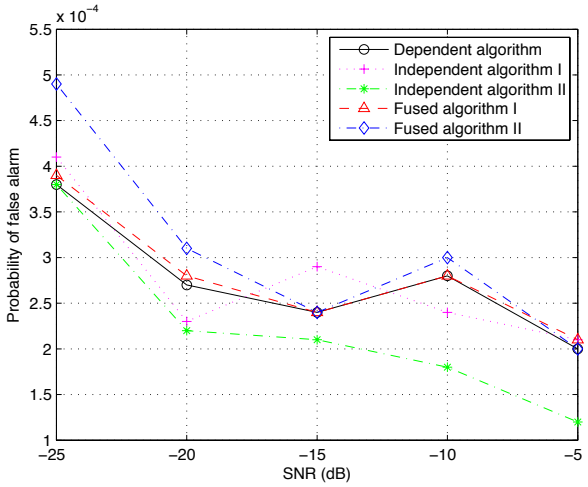

Fig. 16. Probability of False Alarm vs. SNR

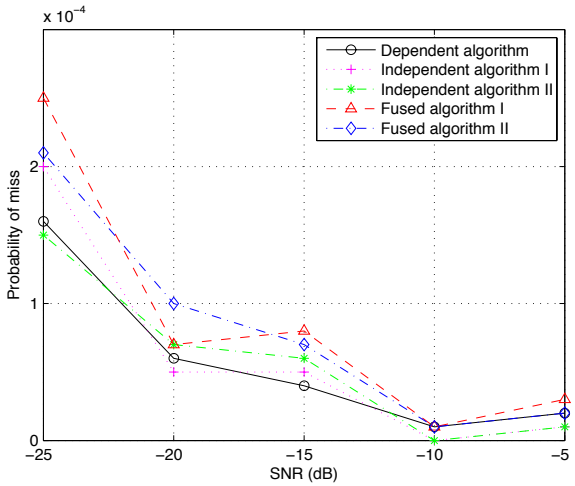

Fig. 17. Probability of Miss vs. SNR

SPRT is weakly admissible [15]. Although the probabilities of error of fused algorithm is higher than the dependent algorithm, the ASN required by the fused algorithm is smaller.

3.4 Conclusion

In this chapter, the terminative joint sequential detection and system state estimation algorithm was proposed to guarantee the Wald's SPRT will eventually terminate with probability one. This is a fused algorithm by using both dependent measurements and independent samples. The relationship between the probabilities of error for the fused algorithm and those for Wald's SPRT with only one test statistic was derived. We also derived the upper bounds on the actual probabilities of error by using nominal probabilities of error. Simulation results showed that the proposed fused terminative algorithm needs a smaller ASN than Wald's SPRT with only one test statistic. The theoretical upper bounds on the actual probabilities of error were demonstrated to be true by simulation results. The theoretical relationship between the proposed fused terminative algorithm and Wald's SPRT with only one test statistic was also demonstrated to be true by simulations.

CHAPTER 4

JOINT GROUP TESTING OF TIME-VARYING FAULTY SENSORS AND SYSTEM STATE ESTIMATION

4.1 Introduction

Faulty sensors are the sensors which return corrupted data [39, 40, 41]. The corruption could be caused by attacks from an adversary [42, 43, 44], sensor malfunctioning [45], or disturbance from the environment [46]. One interesting topic is how to reliably estimate system state when faulty sensors exist in large sensor networks.

In many cases, the faulty sensors are sparse in sensor networks. For example, if sensors are attacked by an adversary, typically only a small number of sensors are attacked and corrupted due to the adversary's limited resources and his/her intention to reduce the chance of being detected by the system defender. Furthermore, the adversary may adopt a time-varying attack strategy to further reduce the probability of being detected. In a real environment, the measurement noise of closely located sensors is usually correlated. In this case, detecting faulty sensors by testing each single sensor individually is not optimal. One promising approach to solve these problems is group testing which is able to detect sparse faulty sensors using testing groups. It was first proposed by Dorfman [47] to identify infected soldiers by detecting syphilitic antigen in blood sample pool during the World War II. In group testing, each test is applied to a subset of populations instead of performing tests on every single item separately. The subsets of a population is called testing groups (or testing pools). Group testing can be viewed as a Boolean version of compressive sensing [48, 49, 50], where a sparse vector only consists of binary entries and Boolean matrix

multiplication is used to generate compressed testing results which contain all the information of the sparse vector.

There are few papers on fault/failure detection using group testing in dynamic systems with time-varying fault states. One related publication is [51], in which a fault detection approach based on combinatorial group testing and the Kalman filter was proposed. In this approach, each testing group is divided into two subgroups and two Kalman filters are run separately on them. The decision for each testing group is made by comparing predicted state estimates of the two Kalman filters. Note that in [51], only the problem of *time-invariant* faulty sensor detection was investigated. In [52], the problem of sparse fault/failure detection in distributed sensor networks was studied. To reduce communication cost, the group testing procedure was successively separated into two phases, in which all the sensors only need to communicate with their neighbors. However, this approach requires the fault state to be *time-invariant* while preforming group testing over phases. In addition, both the above mentioned approaches perform only sensor fault/failure detection but not system state estimation.

Typically, faulty sensor detection and system state estimation are performed separately and the latter is implemented after all the faulty sensors are detected and removed from the system. To detect the faulty sensor(s) via group testing, a time consuming optimization problem needs to be solved. Therefore, it is difficult to implement this procedure in real-time systems. In this chapter, an approach for joint multiple time frame group testing of time-varying faulty sensors and system state estimation is proposed, in which system state is estimated in real time before decoding the fault state of sensors. Furthermore, the sensors are tested in groups which will improve detection performance when sensors' measurement noise is correlated. The innovation of Kalman filter is used to decide whether a testing group contains faulty

sensors or not.

This chapter is organized as follows. The problem is formulated in Section 4.2. The multiple time frame group testing method is proposed in Section 4.3. The joint group testing of time-varying faulty sensors and system state estimation is proposed in Section 4.4. In Section 4.5, several simulations are implemented to show the performance of proposed approach. Finally, this chapter is concluded in Section 4.6.

4.2 Problem Formulation

In this chapter, we still assume that the system state is modeled by the following discrete-time linear system state equation [4]

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{v}_k \tag{4.1}$$

where $\mathbf{x}_k$ is the $n_x \times 1$ state vector at time k, $\mathbf{F}$ is the $n_x \times n_x$ state transition matrix, and $\{\mathbf{v}_k\}$ is a sequence of white Gaussian process noise with $E\{\mathbf{v}_k\} = \mathbf{0}$ and $E\{\mathbf{v}_k\mathbf{v}_k^T\} = \mathbf{Q}_k$ for all $k = 0, 1, 2, \ldots$.

Let us consider a large sensor network which is composed of N sensors. Denote this sensor network as a set $\mathcal{N} = \{1, 2, \ldots, N\}$. Assume that only a few sensors in the sensor network are corrupted by adversary and the fault state of sensors is time-varying. Denote the set of faulty sensors at time k as $\mathcal{D}_k$ which is a subset of $\mathcal{N}$. The components of $\mathcal{D}_k$ are time-varying as different sensors are attacked over time. Denote the size of $\mathcal{D}_k$ by D_k, which is also time-varying and $D_k \ll N$. To detect faulty sensors, the possible states of each sensor are represented by two hypotheses H_0 and H_1 respectively. Let us assume that under hypothesis H_0, sensor i is normal

and its measurement equation is

$$\mathbf{z}_k^i = \mathbf{H}^i \mathbf{x}_k + \mathbf{w}_k^i \tag{4.2}$$

where $\mathbf{z}_k^i$ is the $n_z \times 1$ measurement vector of sensor i at time k, $\mathbf{H}^i$ is the $n_z \times n_x$ measurement matrix of sensor i, and $\mathbf{w}_k^i$ is the measurement noise of sensor i at time k. Also, $\{\mathbf{w}_k^i\}$ is a sequence of white Gaussian measurement noise with $E\{\mathbf{w}_k^i\} = \mathbf{0}$ for $k = 1, 2, \ldots$ and $i = 1, 2, \ldots, N$. The measurement noise of different sensors is assumed to be correlated.

Under hypothesis H_1, sensor i is faulty and its measurement equation is

$$\mathbf{z}_k^i = \mathbf{H}^i \mathbf{x}_k + \mathbf{w}_k^i + \mathbf{b}_k^i \tag{4.3}$$

where $\mathbf{b}_k^i$ is the bias vector which is injected by the adversary to sensor i at time k.

The Kalman filter is used to estimate the system state and generate the innovation (measurement residual) which is used to detect the faulty sensors. Considering the measurement noise of the sensors is correlated, block processing is used to update state estimate when using Kalman filter, which means that state estimate and state covariance are updated at time k using the sensor measurement vector consisting of the measurements from all the sensors, i.e., $\mathbf{z}_k = \begin{bmatrix} (\mathbf{z}_k^1)^T & (\mathbf{z}_k^2)^T & \cdots (\mathbf{z}_k^N)^T \end{bmatrix}^T$ [4]. To maintain the performance of the Kalman filter in the presence of faulty sensors, the measurements of time-varying faulty sensors should be removed adaptively. This motivates joint group testing of time-varying faulty sensors and system state estimation.

4.3 Multiple Time Frame Group Testing

Since a large sensor network is considered and the faulty sensors are assumed to be sparse in the sensor network, group testing is adopted to detect sensor faults.

Group testing implements tests on several testing groups which are generated by binary probabilistic sampling matrix, and the indicator vector of defective sensors is decoded from the testing results. Typically, group testing is applied at each point in time. In this chapter, a new group testing structure is designed over a period of time. By doing this, we are able to detect faulty sensors when their quantity and indices are changing over time.

The fault state of all the N sensors during a time period K is indicated by a KN-dimensional binary vector $\mathbf{f} \in \mathrm{GF}^{KN}(2)$, where GF(2) is a Galois field of order two [53]. $\mathbf{f}(i) = 1$ indicates sensor $1+[(i-1) \bmod N]$ at time $\lceil i/N \rceil$ is faulty whereas $\mathbf{f}(i) = 0$ indicates a normal sensor. Denote the sparsity level of $\mathbf{f}$ by d, and clearly $d = \sum_{k=1}^{K} D_k$. Assume T_g testing groups are generated in total. The tests performed on the sensor network over sensors and over time are represented with $T_g \times KN$ probabilistic sampling matrix $\mathbf{\Phi}$. If $\mathbf{\Phi}(i,j) = 1$, then the sensor $1 + [(j-1) \bmod N]$ at time $\lceil j/N \rceil$ is selected in the ith testing group. The entries of $\mathbf{\Phi}$ follow i.i.d. Bernoulli(p). In noise-free model, group testing outcome vector $\mathbf{g}$ is obtained as follows

$$\mathbf{g} = \mathbf{\Phi} \odot \mathbf{f} \tag{4.4}$$

where $\mathbf{g} \in \mathrm{GF}^{T_g}(2)$, and $\odot$ denotes the Boolean matrix multiplication operator which is composed of the logical AND and OR operators. The tests are called positive tests if the corresponding outcomes are ones; otherwise, the tests are negative tests. In the presence of noise, group testing results are inverted which can be illustrated by the following simple model [54]

$$\mathbf{g} = (\mathbf{\Phi} \odot \mathbf{f}) \oplus \mathbf{e} \tag{4.5}$$

where $\oplus$ denotes XOR operator, $\mathbf{e} \in \mathrm{GF}^{T_g}(2)$ is the Boolean vector of errors which represents the effect of noise. The ones in $\mathbf{e}$ indicate corruption and they will invert

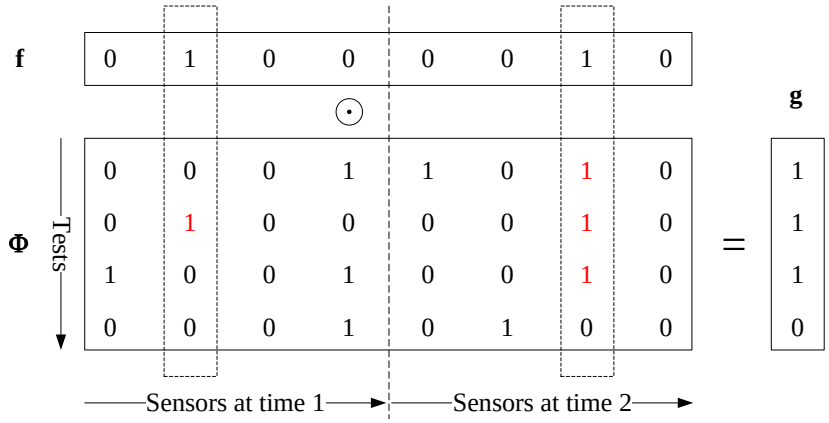

Fig. 18. An example of time-varying group testing problem

the corresponding results of $\mathbf{\Phi} \odot \mathbf{f}$ which leads to false alarms or misses. Note that model (4.5) is a simple way to illustrate the presence of noise/errors and $\mathbf{e}$ is difficult to model statistically in some cases.

A toy example of the noise-free multiple time frame group testing procedure is shown in Fig. 18. In this example, $N = 4$, $K = 2$, $T_g = 4$, and $D_1 = D_2 = 1$. Sensor 2 at time 1 and Sensor 3 at time 2 are faulty. Sensor 2 at time 1 is selected in Test 2 and Sensor 3 at time 2 is selected in Tests 1, 2, and 3. As long as at least one faulty sensor is selected in a specific test, the outcome of this test will be 1. If no faulty sensors are selected in one test, then its testing outcome is 0. Therefore, Tests 1, 2, and 3 are positive tests and their outcomes $g_1 = g_2 = g_3 = 1$, Test 4 is a negative test and its outcome $g_4 = 0$.

To decode the fault state vector $\mathbf{f}$ efficiently, the probabilistic sampling matrix $\mathbf{\Phi}$ should satisfy d-disjunct property as it ensures identifiability of the d-sparse fault state vector. A matrix $\mathbf{\Phi}$ is called d-disjunct if for any $d + 1$ columns, there always exists a row with entry 1 in a column and zeros in all the other d columns [55]. In the example shown in Fig. 18, $\mathbf{\Phi}$ is a 2-disjunct probabilistic sampling matrix. The fault

state vector $\mathbf{f}$ is decoded via the linear programming (LP) relaxation [54], in which the inputs are $\mathbf{g}$ and $\mathbf{\Phi}$, and the output is $\mathbf{f}$.

$$\begin{aligned} &\min \sum_{s=1}^{KN} f_s + \rho \sum_{t=1}^{T_g} \xi_t \\ &\text{s.t. } \mathbf{0} \leqslant \mathbf{f} \leqslant \mathbf{1},\ \mathbf{0} \leqslant \boldsymbol{\xi}_{\mathcal{I}} \leqslant \mathbf{1},\ \boldsymbol{\xi}_{\mathcal{J}} \geqslant \mathbf{0} \\ &\mathbf{\Phi}_{\mathcal{I}}\mathbf{f} + \boldsymbol{\xi}_{\mathcal{I}} \geqslant \mathbf{1},\ \mathbf{\Phi}_{\mathcal{J}}\mathbf{f} = \boldsymbol{\xi}_{\mathcal{J}} \end{aligned} \tag{4.6}$$

where f_s denotes the sth entry of $\mathbf{f}$ and ρ is a regularization parameter. $\mathcal{I} = \{i|g_i = 1\}$ denotes the set of positive tests, $\mathcal{J} = \{j|g_j = 0\}$ denotes the set of negative tests. $\boldsymbol{\xi} = \boldsymbol{\xi}_{\mathcal{I}} \cup \boldsymbol{\xi}_{\mathcal{J}}$ is a set of slack variables where $\boldsymbol{\xi}_{\mathcal{I}}$ and $\boldsymbol{\xi}_{\mathcal{J}}$ correct false alarms and misses caused by $\mathbf{e}$ in (4.5), respectively. Denote the tests with false alarm outcomes as $\mathcal{I}_f$. The false alarms $g_{\mathcal{I}_f}$ can be corrected by introducing the fourth constraint in (4.6). Similarly, the misses $g_{\mathcal{I}_m}$ can be corrected by introducing the fifth constraint in (4.6). According to [54], this LP relaxation provides the optimal solution if $\mathbf{f}$ is d-sparse and $\mathbf{\Phi}$ is d-disjunct. Some algorithms, such as the ellipsoid algorithm and Karmarkar's algorithm, can solve this problem in polynomial time even in the worst case scenarios.

4.4 Joint Group Testing of Time-varying Faulty Sensors and System State Estimation

Note that the outcome of group testing is a binary vector but the measurements are continuous. We need to find a way to decide whether a testing group contains faulty sensors or not. The innovation of Kalman filter is a good choice as it is a zero-mean, white, and Gaussian sequence when no faulty sensors are in the testing group. To achieve real-time system state estimation and detect time-varying faulty sensors, joint group testing of time-varying faulty sensors and system state estimation

is proposed and described as the following steps.

Step I: Build testing groups. Generate $T_g \times KN$ probabilistic sampling matrix $\mathbf{\Phi}$ via Bernoulli(p). Let us divide $\mathbf{\Phi}$ into K blocks by column, where each block is a $T_g \times N$ sub-matrix $\mathbf{\Phi}_k$. Denote the t-th testing group at time k by $\mathcal{G}_{t,k}$, which is chosen by the t-th row of $\mathbf{\Phi}_k$. The size of $\mathcal{G}_{t,k}$ is denoted by $G_{t,k}$, where $1 \leqslant t \leqslant T_g$ and $1 \leqslant k \leqslant K$.

Step II: Generate outcome vector $\mathbf{g}$. Run Kalman filter from $k = 1$ to K. For each time k, run Kalman filter by using each testing group $\mathcal{G}_{t,k}$, then obtain innovation $\boldsymbol{\nu}_{t,k}$ and its covariance $\mathbf{S}_{t,k}$. If all the sensors in $\mathcal{G}_{t,k}$ are normal, then $\boldsymbol{\nu}_{t,k}$ is a zero-mean Gaussian random variable and it can be tested via χ^2 test: $\boldsymbol{\nu}_{t,k}^T \mathbf{S}_{t,k}^{-1} \boldsymbol{\nu}_{t,k} \sim \chi^2(n_z G_{t,k})$. Moreover, the innovation is a zero-mean white sequence if there are no faulty sensors in $\mathcal{G}_{t,1}, \mathcal{G}_{t,2}, \ldots, \mathcal{G}_{t,k}$, and the following holds

$$\sum_{s=1}^{k} \boldsymbol{\nu}_{t,s}^T \mathbf{S}_{t,s}^{-1} \boldsymbol{\nu}_{t,s} \sim \chi^2 \left(n_z \sum_{s=1}^{k} G_{t,s} \right) \tag{4.7}$$

Note that if $\mathcal{G}_{t,k} = \emptyset$, do not run Kalman filter in test t at time k and skip the corresponding item in (4.7). The outcome vector $\mathbf{g}$ is generated via (4.7) as follows: the testing groups $\mathcal{G}_{t,1}, \mathcal{G}_{t,2}, \ldots, \mathcal{G}_{t,K}$ in test t are tested one by one from $k = 1$ to $k = K$. If (4.7) is unsatisfied at time k, this procedure is stopped for test t with positive outcome, i.e. $g_t = 1$. Otherwise, the next innovation $\boldsymbol{\nu}_{t,k+1}$ is calculated and tested. If (4.7) is satisfied by all the testing groups in test t, the outcome of test t is negative and $g_t = 0$. In this way, not all testing groups chosen by $\mathbf{\Phi}$ are fully tested as this procedure may stop when $k < K$, which saves computational costs.

Step III: State estimation via Kalman filter. At each time k, test all the sensor groups $\mathcal{G}_{1,k}, \mathcal{G}_{2,k}, \ldots, \mathcal{G}_{T_g,k}$ via (4.7). Form a normal sensor group by taking the union of all the sensor groups which pass the test. Run Kalman filter on this normal sensor

group using block processing. The updated state estimate and covariance are used as the inputs of Kalman filter at the next time $k+1$ in both Steps II and III, and they are the system state estimation outputs of this algorithm at time k.

Step IV: Identify faulty sensors via group testing. By the end of time K, with the group testing outcome vector $\mathbf{g}$, the fault state vector $\mathbf{f}$ is decoded by solving (4.6).

Note that system state estimation is implemented before decoding $\mathbf{f}$ which is time consuming. This design guarantees real-time system state estimation with faulty sensors in large sensor networks. If only system state estimation is required, then Step IV does not need to be implemented.

If the sensors are tested via χ^2 test one by one at each time, we can design a similar testing procedure. The differences are $T_g = N$, $\mathbf{\Phi}_k = \mathbf{I}$ for all $k \in \{1, 2, \ldots, K\}$, and there is no need to calculate group testing outcome vector $\mathbf{g}$. Now we analyze the computational complexities of group testing based approach and one-by-one test approach. We assume that the adversary chooses sensors to attack randomly via Bernoulli($q(K, N)$). We use q instead of $q(K, N)$ for simplicity. For both approaches, the dominating term in computational complexity is for the inverse of measurement prediction covariance in Step III, which is $\mathcal{O}\left(\left\{\left[1-(1-p)^{T_g}\right](1-q)Nn_z\right\}^3\right)$ for group testing based approach and $\mathcal{O}\left([(1-q)Nn_z]^3\right)$ for the one-by-one test approach. Since Step III will be implemented K times, the computational complexity of group testing based approach is $\mathcal{O}\left(K\left\{\left[1-(1-p)^{T_g}\right](1-q)Nn_z\right\}^3\right)$. For the one-by-one test approach, the computational complexity is $\mathcal{O}\left(K[(1-q)Nn_z]^3\right)$. We know that $p = \frac{1}{qKN}$ design is asymptotically close to optimal when $qKN = o(KN)$ [55, 56]. If $T_g = \mathcal{O}(qKNlog(KN))$, the randomly generated matrix $\mathbf{\Phi}$ will be qKN-disjunct with an arbitrarily small error probability [57]. Hence, $\left[1-(1-p)^{T_g}\right]^3 \to 1$ when $KN \to \infty$. Therefore, the computational complexity of the proposed group testing

based method and that of the one-by-one test based method are at the same level.

Table 4. Computational complexity

Group testing based approach	One-by-one test
$\mathcal{O}\left(K\left\{\left[1-(1-p)^{T_g}\right](1-q)Nn_z\right\}^3\right)$	$\mathcal{O}\left(K[(1-q)Nn_z]^3\right)$

4.5 Simulation Results

For simplicity we give a multi-sensor target tracking example to illustrate the effectiveness of the proposed approach. Let us assume that an object is moving in a 1-dimensional space with its state at time k denoted by $\mathbf{x}_k = [\varphi_k \quad \dot{\varphi}_k]^T$, where φ_k and $\dot{\varphi}_k$ are the object's position and velocity at time k, respectively. The state transition matrix is $\mathbf{F} = [1 \ T_s;\ 0 \ 1]$ where $T_s = 0.1$ seconds is the time interval between two measurements. The covariance matrix of state process noise is $\mathbf{Q}_k = 0.01 \times \left[\frac{T_s^4}{4} \ \frac{T_s^3}{2};\ \frac{T_s^3}{2} \ T_s^2\right]$ for all k. The mean and covariance matrix of the object's initial state are $\hat{\mathbf{x}}_{0|0} = [0 \ \ 1.5]^T$ and $\mathbf{P}_{0|0} = diag([1000, 1])$, respectively. Assume that there is a linear sensor array includes $N = 150$ sensors, the distance between adjacent sensors is 2m. All the sensors measure the object's position over time. Namely, the measurement matrix is $\mathbf{H}^i = [1 \ \ 0]$ for all i. The element of covariance matrix of sensor measurement noise in the ith row and the jth column is $\mathbf{R}_w(i,j) = \sigma_w^2 e^{-\rho\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|_2}$ [58], where $\sigma_w = 1$, $\rho = 0.01$, $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_j$ are positions of sensor i and sensor j respectively. The adversary chooses sensors to attack randomly via Bernoulli(q) where $q = 0.03$. The bias injected by the adversary follows i.i.d. Gaussian distribution $b_k^i \sim \mathcal{N}(0, R_b)$. We choose $K = 5$ to design $\mathbf{\Phi}$ and the entries of $\mathbf{\Phi}$ follow i.i.d. Bernoulli(p), where $p = \frac{1}{qKN} = 0.0444$. The number of testing groups is $T_g = 200$. Two-sided χ^2 test with 0.005 significance level is applied in Step II in Section 4.4. The regularization parameter ρ in (4.6) is set as 1. All the results are based on 100 Monte Carlo runs.

The tracking performance of the proposed approach is compared with two approaches: one is using the measurements from all the sensors, and the other is the one-by-one test based approach. The performance of the three approaches is compared in terms of root mean squared error (RMSE) of the position estimate. We assume that $R_b = 5 \times 10^4$ which means that the injected bias noise by the adversary is strong. The simulation results are shown in Fig. 19. It is clear that the RMSEs of the position estimate of the proposed approach are the smallest among the three approaches and they are almost the same as the RMSEs achieved by a clairvoyant Kalman filter using all the normal sensors. That is to say, the proposed approach chooses normal sensors efficiently when tracking the object. One reason is that in group testing, there are multiple sensors in each testing group, and the χ^2 test can take advantage of the correlation between sensor measurements. Furthermore, the test statistic in group testing (shown in (4.7)) has higher degrees of freedom and less variability than that used in one-by-one test. From Fig. 19, it is clear that the proposed group testing based approach is very robust to attacks with strong injection noise.

To study the fault detection performance of the proposed approach under different levels of attacks, R_b is changed from 10^3 to 5×10^4 and probabilities of error are evaluated under different R_b. The simulation results are shown in Fig. 20. We can see that both the probabilities of miss and false alarm of group testing based approach are less than those of the one-by-one test based approach. Again, the reason is that group testing based approach takes advantage of the correlated measurements from multiple sensors when detecting faulty sensor(s), whereas the one-by-one test has to make a decision relying on a single sensor's measurement.

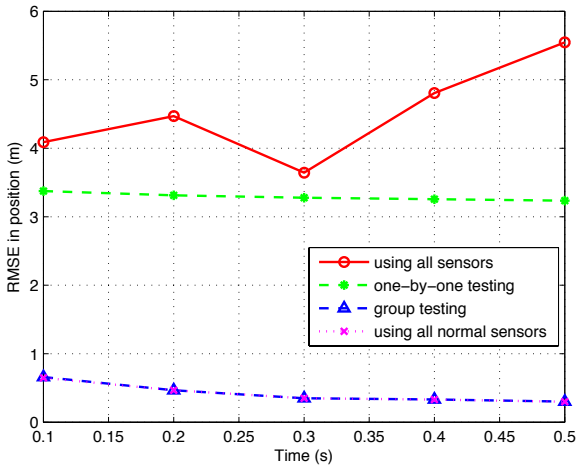

Fig. 19. RMSE of the position estimate over time

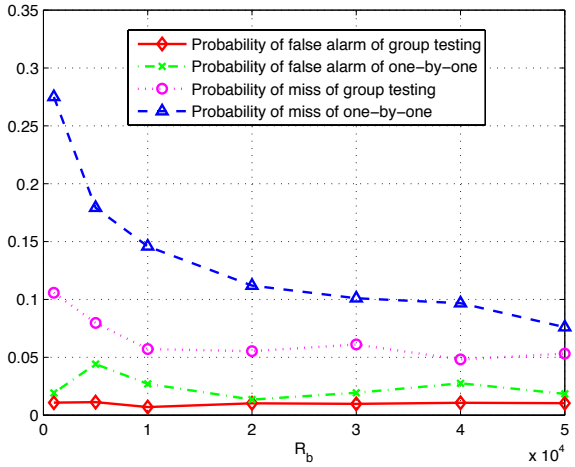

Fig. 20. probabilities of error under different $\mathbf{R}_b$

4.6 Conclusion

In this chapter, an approach for joint group testing of time-varying faulty sensors and system state estimation was proposed. A multiple time frame group testing structure was developed to detect the time-varying fault state of sensors. For real time implementation, system state estimation is performed without waiting for the

estimate of the full fault state of sensors. The proposed approach has the same level of computational complexity as an alternative one-by-one test. Simulations showed that, compared to the one-by-one test, the proposed approach has significantly better system state estimation performance and improves faulty sensor detection performance, especially the probability of miss. The reason is that the proposed approach uses correlated measurements from multiple sensors to make decision and the degrees of freedom of the χ^2 test statistics are higher with less variability. Furthermore, the proposed approach achieves almost the same state estimation performance as a clairvoyant Kalman filter with the perfect knowledge of the sensor fault state.

CHAPTER 5

MINIMAX ANTI-JAMMER DESIGN FOR FHSS/QPSK SATELLITE COMMUNICATION SYSTEMS

5.1 Introduction

Jamming attacks are significant threats to satellite communication systems. A jammer can simply interfere the legitimate communication systems by injecting jamming signals into communication channels [1, 2]. As a result, the communication between the transmitter and receiver will be corrupted. To achieve reliable communication, employing anti-jammer is crucial in satellite communication systems. One efficient way to design anti-jammer is adopting frequency hopping spread spectrum (FHSS) [59, 60], which avoids attacks by switching channels from time to time.

In the literature, the performance of phase-shift keying (PSK) with FHSS was studied under various types of jammer attacks. In [61], signals were generated using binary PSK (BPSK) with FHSS. Four types of jamming were investigated and the bit error probability was provided under each type of jamming. In [62], the error probability of quadrature PSK (QPSK) or quadrature amplitude-shift keying (QASK) with FHSS was evaluated, either in the presence of partial-band multitone jamming or partial-band noise jamming. In [63], the performance of a slow frequency-hopped differential PSK system was studied in the presence of jamming and additive white Gaussian noise (AWGN). The performance of any M-ary PSK with FHSS under partial-band multitone jamming was studied in [64]. By assuming that the phase of jamming signal is a random variable, the probability density function (PDF) of the phase difference between received jammed signal and transmitted signal was derived,

based on which the bit error probability and the performance expression under worst jamming scenarios were obtained in [64]. In all the above publications, no anti-jammer was considered and the decision thresholds used by the demodulator are fixed. An optimal anti-jammer was designed by solving a maximin optimization problem in [65]. In this maximin optimization problem, the expected detection time at the jammer was minimized respect to the window length used by the jammer's detector and maximized respect to the receiver side signal power.

In this chapter, we design an anti-jammer which is optimal under the worst jamming scenario. The FHSS/QPSK is employed to modulate signals. Since the main purpose of satellite communication systems is to transfer signals correctly, the performance of FHSS/QPSK is evaluated by using the symbol error probability. The purpose of the anti-jammer is to minimize the symbol error probability while that of the jammer is to maximize it. To reduce the jammer's effect, the anti-jammer uses FHSS to transmit and receive QPSK signals. The carrier frequency of QPSK signals is hopping between N frequencies according to a predetermined pseudo-random frequency hopping sequence. We assume that the jammer uses the same symbol duration as the transmitter and receiver, and the total jamming power is fixed. For simplicity, we assume that the jammer attacks the legitimate communication system with only two jamming strategies: single-tone jamming—randomly jamming one tone with full power, and full-band multitone jamming—jamming all the N tones and allocating jamming power equally among the N tones. The jamming tones coincide with carrier frequencies of the transmitter and receiver. In the worst jamming case, jammer will use the strategy and jamming phase which cause highest symbol error probability. Therefore, the anti-jammer is designed to minimize symbol error probability under the worst jamming case by changing decision thresholds when the receiver demodulates QPSK signals. So, the anti-jammer needs to solve a minimax optimization

problem.

This chapter is organized as follows. The problem is formulated in Section 5.2. The minimax anti-jammer is design in Section 5.3. In Section 5.4, several numerical results are shown to reveal the properties of the minimax anti-jammer. Finally, this chapter is concluded in Section 5.5.

5.2 Problem Formulation

In this chapter, QPSK is used to modulate and demodulate the signals in satellite communication systems. Let $U_k(t)$ be the QPSK signal sent by transmitter, where k is signal index. Assume the jammer attacks the communication systems by jamming interference $J_k(t)$ into the channel. Denote the received signal by $R_k(t)$, which is represented as

$$R_k(t) = U_k(t) + J_k(t) + n(t),\ k = 1, 2, \ldots \tag{5.1}$$

where $n(t)$ denotes the sample function of the additive white Gaussian noise (AWGN) random process with the power spectral density $S_n(f) = \frac{N_0}{2}$ W/Hz.

The QPSK modulated signal $U_k(t)$ is a frequency hopping signal, which is represented as

$$U_k(t) = p_k(t)\cos(2\pi f_c(k)t + \phi_m(k)) \tag{5.2}$$

where $f_c(k)$ is the carrier frequency for the kth signal which is changing among N tones $\{f_{c1}, f_{c2}, \ldots, f_{cN}\}$ according to a predetermined pseudo-random frequency hopping sequence. The phase for the kth signal is

$$\phi_m(k) = \frac{(2m-1)\pi}{4},\ m = 1, 2, 3, 4 \tag{5.3}$$

and $p_k(t)$ is

$$p_k(t) = \sqrt{\frac{2E_s}{T_m}} g_{T_m}[t - (k-1)T_m] \tag{5.4}$$

where E_s is the energy in each signal, T_m is the signal duration, and $g_{T_m}(t)$ is a rectangular baseband pulse as follows

$$g_{T_m}(t) = \begin{cases} 1, & 0 \leqslant t < T_m \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

Two jamming strategies are considered: single-tone jamming and full-band multitone jamming. The single-tone jamming is randomly jamming one tone with full power at each time. The full-band multitone jamming is jamming all the N tones with equal power. Assuming that the jammer uses the same signal duration T_m, the jamming signal is in one of the following forms

$$\begin{aligned} &\text{Strategy I}: \quad J_k^{\mathrm{I}}(t) = q_k(t)\cos(2\pi f_j(k)t + \theta) \\ &\text{Strategy II}: \quad J_k^{\mathrm{II}}(t) = \textstyle\sum_{i=1}^{N} \frac{q_k(t)}{\sqrt{N}} \cos(2\pi f_{ci}t + \theta) \end{aligned} \tag{5.6}$$

where $f_j(k)$ is carrier frequency of the kth jamming signal which is randomly picked from the N tones $\{f_{c1}, f_{c2}, \ldots, f_{cN}\}$, the phase θ is deterministic which is chosen from $[0, 2\pi)$, and $q_k(t)$ is

$$q_k(t) = \sqrt{\frac{2E_j}{T_m}} g_{T_m}[t - (k-1)T_m] \tag{5.7}$$

where E_j is the energy in each jamming signal.

The orthonormal basis functions used by the receiver are

$$\begin{cases} \psi_{1k}(t) = \sqrt{\frac{1}{E_s}} p_k(t)\cos(2\pi f_c(k)t) \\ \psi_{2k}(t) = -\sqrt{\frac{1}{E_s}} p_k(t)\sin(2\pi f_c(k)t) \end{cases} \tag{5.8}$$

The outputs of QPSK demodulator $\mathbf{y}_m(k)$ contains two signal components $y_{1m}(k)$ and $y_{2m}(k)$, which may be expressed as

$$\mathbf{y}_m(k) = \begin{bmatrix} y_{1m}(k) \\ y_{2m}(k) \end{bmatrix} = \mathbf{s}_m(k) + \mathbf{j}(k) + \mathbf{n}(k) \tag{5.9}$$

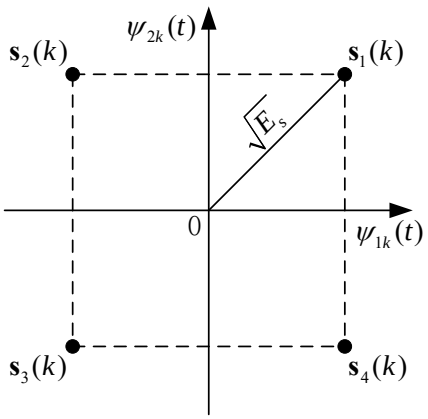

Fig. 21. QPSK signal constellation.

Let $\mathbf{s}_m(k) = \begin{bmatrix} s_{1m}(k) \\ s_{2m}(k) \end{bmatrix}$, $\mathbf{j}(k) = \begin{bmatrix} j_1(k) \\ j_2(k) \end{bmatrix}$, and $\mathbf{n}(k) = \begin{bmatrix} n_1(k) \\ n_2(k) \end{bmatrix}$. The two components of $\mathbf{s}_m(k)$ are

$$\begin{cases} s_{1m}(k) = \sqrt{E_s}\cos(\phi_m(k)) \\ s_{2m}(k) = \sqrt{E_s}\sin(\phi_m(k)) \end{cases} \tag{5.10}$$

From (5.10), we know that when the channel is noise-free and jamming-free, the four symbols $\mathbf{s}_m(k)$ $(m = 1, 2, 3, 4)$ form a square in phase quadrature, which is shown in Fig. 21.

The two components of $\mathbf{n}(k)$ are

$$\begin{cases} n_1(k) = \frac{1}{\sqrt{2T_m}} \int_0^{T_m} n_{ck}(t)dt \\ n_2(k) = \frac{1}{\sqrt{2T_m}} \int_0^{T_m} n_{sk}(t)dt \end{cases} \tag{5.11}$$

where $n_{ck}(t)$ and $n_{sk}(t)$ are the in-phase and quadrature components of $n(t)$ respectively, i.e., $n(t) = n_{ck}(t)\cos(2\pi f_c(k)t) - n_{sk}(t)\sin(2\pi f_c(k)t)$. From (5.11), it is easy to show that $n_1(k)$ and $n_2(k)$ are independent and identically distributed Gaussian random variables with zero mean and variance $\frac{N_0}{2}$. When the channel is jamming-free, the outputs of QPSK demodulator follow Gaussian distribution $\mathbf{y}_m(k) \sim \mathcal{N}(\mathbf{s}_m(k),\ \frac{N_0}{2}\mathbf{I})$.

If Strategy I is used, $\mathbf{j}(k) = \mathbf{0}$ when $f_j(k) \neq f_c(k)$. When $f_j(k) = f_c(k)$, the two components of $\mathbf{j}(k)$ are

$$\begin{cases} j_1^{\mathrm{I}}(k) = \sqrt{E_j}\cos\theta \\ j_2^{\mathrm{I}}(k) = \sqrt{E_j}\sin\theta \end{cases} \tag{5.12}$$

If Strategy II is used, the two components of $\mathbf{j}(k)$ are

$$\begin{cases} j_1^{\mathrm{II}}(k) = \sqrt{\frac{E_j}{N}}\cos\theta \\ j_2^{\mathrm{II}}(k) = \sqrt{\frac{E_j}{N}}\sin\theta \end{cases} \tag{5.13}$$

Obviously, when the received signal contains additive white Gaussian noise and deterministic interference, the outputs of QPSK demodulator follow Gaussian distribution $\mathbf{y}_m(k) \sim \mathcal{N}(\mathbf{s}_m(k) + \mathbf{j}(k),\ \frac{N_0}{2}\mathbf{I})$. To find the best anti-jammer under the worst jamming case, we design the minimax anti-jammer which is shown in the next section.

5.3 Minimax Anti-Jammer Design

Since the symbol error probability is the most important measure to evaluate the performance of the QPSK communication system, the aim of the anti-jammer is to minimize it while the jammer's goal is to maximize it. The best anti-jammer under the worst jamming case can be obtained by solving a minimax optimization problem, in which the objective function is the maximum symbol error probability of the two jamming strategies. Denote symbol error probability by $P_e(\cdot)$. The jammer may change $P_e(\cdot)$ via $\mathbf{j}(k)$, and the anti-jammer may change $P_e(\cdot)$ via decision thresholds. Let ξ_k and η_k denote the decision thresholds along $\psi_{1k}(t)$ and $\psi_{2k}(t)$ respectively. Then, the symbol error probability is a function of ξ_k, η_k, $j_1(k)$, and $j_2(k)$. We omit k for simplicity. Then, symbol error probability is denoted by $P_e(\xi, \eta, j_1, j_2)$.

From (5.6), we know that the jammer may attack the channel via two strategies, and the jamming signal may use different phases for each strategy. When Strategy I

is used, the probability of $f_j(k) = f_c(k)$ is $\frac{1}{N}$, and the probability of $f_j(k) \neq f_c(k)$ is $\frac{N-1}{N}$. The symbol error probability of Strategy I is

$$P_e^{\mathrm{I}} = \frac{1}{N} P_e(\xi, \eta, j_1^{\mathrm{I}}, j_2^{\mathrm{I}}) + \frac{N-1}{N} P_e(\xi, \eta, 0, 0) \tag{5.14}$$

where j_1^{I} and j_2^{I} are provided in (5.12).

When Strategy II is used, the symbol error probability is

$$P_e^{\mathrm{II}} = P_e(\xi, \eta, j_1^{\mathrm{II}}, j_2^{\mathrm{II}}) \tag{5.15}$$

where j_1^{II} and j_2^{II} are provided in (5.13). Note that j_1 and j_2 depend on phase θ for both strategies.

The minimax anti-jammer can be formulated as

$$\min_{\xi,\eta} \max_{i, j_1, j_2} P_e^i \tag{5.16}$$

To solve this minimax optimization problem, we need to derive the closed-form of $P_e(\xi, \eta, j_1, j_2)$. We know that the received signal contains two binary phase modulation signals along in-phase and quadrature carriers respectively. With perfect orthonormal basis as in (5.8), there is no crosstalk or interference between the signals on the in-phase and quadrature carriers [66]. Therefore, if Gray encoding is adopted, the symbol error probability of QPSK has the following relationship with bit error probabilities.

$$P_e(\xi, \eta, j_1, j_2) = 1 - [1 - P_{b\xi}(\xi, j_1)] [1 - P_{b\eta}(\eta, j_2)] \tag{5.17}$$

where $P_{b\xi}(\xi, j_1)$ denotes bit error probability on the in-phase carrier $\psi_{1k}(t)$, $P_{b\eta}(\eta, j_2)$ denotes bit error probability on the quadrature carrier $\psi_{2k}(t)$.

Now, we derive the bit error probability $P_{b\xi}(\xi, j_1)$. Since two bits represent one symbol in QPSK, the energy per bit E_b is one half of E_s. From (5.10), we

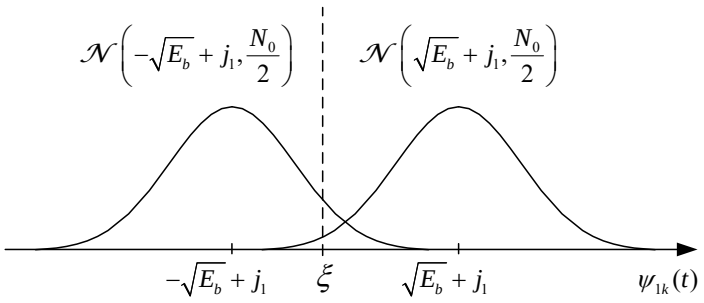

Fig. 22. The PDFs of signals $\mathbf{y}_m(k)$ on the in-phase carrier $\psi_{1k}(t)$.

have $s_{1m}(k) = \sqrt{E_b}$ when $m = 1, 4$, and $s_{1m}(k) = -\sqrt{E_b}$ when $m = 2, 3$. Hence, $y_{1m}(k) \sim \mathcal{N}(\sqrt{E_b} + j_1,\ \frac{N_0}{2})$ when $m = 1, 4$, and $y_{1m}(k) \sim \mathcal{N}(-\sqrt{E_b} + j_1,\ \frac{N_0}{2})$ when $m = 2, 3$. The PDFs of signals $\mathbf{y}_m(k)$ $(m = 1, 2, 3, 4)$ on the in-phase carrier $\psi_{1k}(t)$ are shown in Fig. 22.

From Fig. 22, we can derive the bit error probability $P_{b\xi}(\xi, j_1)$ as follows

$$\begin{aligned} P_{b\xi}(\xi, j_1) &= \frac{1}{2}\int_{-\infty}^{\xi} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(y_a-\sqrt{E_b}-j_1)^2}{N_0}} dy_a \\ &\quad + \frac{1}{2}\int_{\xi}^{+\infty} \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(y_b+\sqrt{E_b}-j_1)^2}{N_0}} dy_b \\ &= \frac{1}{2}\int_{\frac{\sqrt{E_b}+j_1-\xi}{\sqrt{\frac{N_0}{2}}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u_a^2}{2}} du_a \\ &\quad + \frac{1}{2}\int_{\frac{\sqrt{E_b}-j_1+\xi}{\sqrt{\frac{N_0}{2}}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u_b^2}{2}} du_b \end{aligned} \tag{5.18}$$

where $y_a \in \{y_{11}, y_{14}\}$ and $y_b \in \{y_{12}, y_{13}\}$. Using the definition of Q-function $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{+\infty} e^{-\frac{u^2}{2}} du$, the bit error probability $P_{b\xi}(\xi, j_1)$ can be represented as follows

$$P_{b\xi}(\xi, j_1) = \frac{1}{2}Q\left(\frac{\sqrt{E_b}+j_1-\xi}{\sqrt{\frac{N_0}{2}}}\right) + \frac{1}{2}Q\left(\frac{\sqrt{E_b}-j_1+\xi}{\sqrt{\frac{N_0}{2}}}\right) \tag{5.19}$$

By using the same method, it is easy to show that the bit error probability

$P_{b\eta}(\eta, j_2)$ is as follows

$$P_{b\eta}(\eta, j_2) = \frac{1}{2}Q\left(\frac{\sqrt{E_b} + j_2 - \eta}{\sqrt{\frac{N_0}{2}}}\right) + \frac{1}{2}Q\left(\frac{\sqrt{E_b} - j_2 + \eta}{\sqrt{\frac{N_0}{2}}}\right) \tag{5.20}$$

Substituting (5.19) and (5.20) into (5.17), we have the closed-form of $P_e(\xi, \eta, j_1, j_2)$. Then, P_e^{I} and P_e^{II} will be obtained by substituting (5.17) into (5.14) and (5.15). By solving the minimax optimization problem in (5.16), we will have the optimal anti-jammer under the worst jamming attack.

5.4 Numerical Results

In this numerical example, we solve the minimax problem in (5.16). By solving this problem, we will know what the optimal anti-jammer is when it is under the worst case jamming attack, and which strategy and phase the jammer will use if it knows the anti-jammer's decision thresholds. Since j_1 and j_2 are the only functions of θ, we maximize P_e^i respect to θ instead of j_1 and j_2, where $\theta = \frac{l\pi}{32}$, $l = 0, 1, 2, \ldots, 31$. The starting guess of the optimal solution is $\xi = \eta = 0$.

Let $\frac{E_s}{E_j} = 0.5$, $\frac{E_s}{N_0} = 20$, and $N = 5$. By solving (5.16), we have the following results: the optimal solution is $\xi = \eta = 0$ and $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, the optimum objective value is $P_e = 0.1592$. When the anti-jammer has no knowledge about jamming phase θ, no matter which strategy the jammer uses, the optimum decision thresholds for anti-jammer are $\xi = \eta = 0$. If the jammer knows how the anti-jammer designs decision thresholds, the best choice of the jammer under this setting is using Strategy II and $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. To get this optimal solution, the jammer needs to know $\frac{E_s}{E_j}$, $\frac{E_j}{N_0}$, and the anti-jammer's decision rule. Note that (5.16) has more than one optimal solution. To show the optimum solution for the jammer, we plot the optimum objective values of P_e^i with different jamming phases in Fig. 23. Obviously, the optimal solution of

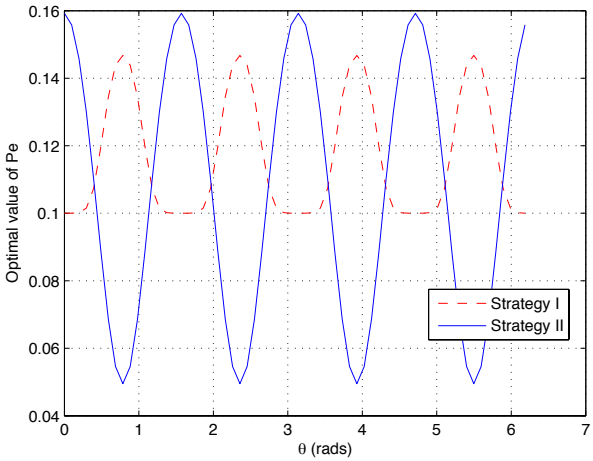

Fig. 23. Optimum objective value of P_e for two strategies with different jamming phases, where $\frac{E_s}{E_j} = 0.5$, $\frac{E_s}{N_0} = 20$, and $N = 5$.

P_e^i are periodic functions of θ with period $\frac{\pi}{2}$. Because the QPSK signal constellation is symmetric, which is shown in Fig. 21, the effect of $\mathbf{j}(k)$ is periodic with period $\frac{\pi}{2}$. More specifically, P_e^{I} is maximized when $\theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$ and P_e^{II} is maximized when $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. It is easy to explain for Strategy II. When $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ and decision thresholds are fixed, $\mathbf{j}(k)$ on the in-phase or quadrature carrier is the largest, which can be figure out from Fig. 22. Hence, the symbol error probability is the largest when $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Fig. 23 also shows that $\max P_e^{\mathrm{I}} < \max P_e^{\mathrm{II}}$ at the optimal solution, which means that jammer can cause more serious interference by applying Strategy II under this setting.

Now, we increase $\frac{E_s}{E_j}$ to 1.25 and show the optimal solution under this setting. By solving (5.16), we have the following results: the optimal solution is $\xi = \eta = 0$ and $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, the optimum objective value is $P_e = 0.0882$. If the jammer knows how the anti-jammer obtains decision thresholds, the best strategy for the jammer becomes Strategy I under this setting. The optimum objective values of P_e^i with different jamming phases are shown in Fig. 24. Obviously, both P_e^{I} and P_e^{II} are

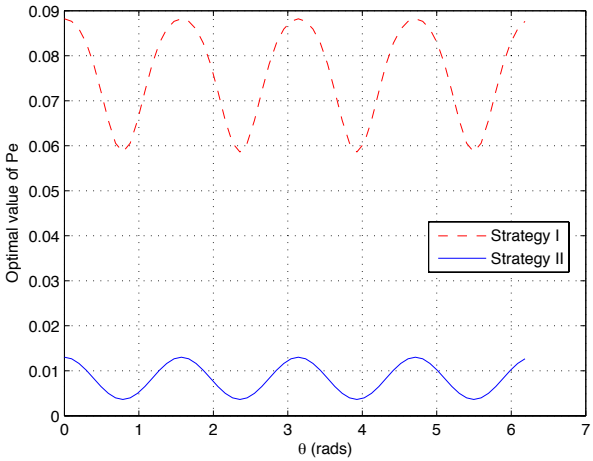

Fig. 24. Optimum objective value of P_e for two strategies with different jamming phases, where $\frac{E_s}{E_j} = 1.25$, $\frac{E_s}{N_0} = 20$, and $N = 5$.

maximized when $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Comparing Fig. 24 with Fig. 23, the maximum positions of P_e^{I} shift $\frac{\pi}{4}$. More specifically, if the working frequency is attacked, $\theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$ will be the optimal solution when the jammer's power is relatively high comparing with the anti-jammer's power; otherwise, $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ will be the optimal solution. We can explain it as follows: when $\frac{E_s}{E_j} = 1.25$, $\sqrt{\frac{E_j}{E_b}} \approx 1.26$ which is not large enough, the jammer needs to fully utilize the limited power by making $\mathbf{j}(k)$ on the in-phase or quadrature carrier the largest, so $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ is the optimal solution; when $\frac{E_s}{E_j} = 0.5$, $\sqrt{\frac{E_j}{E_b}} = 2$, the projection of $\mathbf{j}(k)$ on the in-phase or quadrature carrier is smallest when $\theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$, which is about 1.41 and it is still large, then $\theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$ becomes the best solution for the jammer.

To analyze why the optimal solution is $\xi = \eta = 0$ and $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, we show the maximum P_e between two strategies in ξ-η-θ space in Figs. 25, 26, and 27. First, the function needs to be maximized respect to θ. From Figs. 25 and 26, we know that this function is maximized when θ is close to $\frac{n\pi}{2}, n = 0, 1, 2, 3$. Second, the function needs to be minimized respect to ξ and η. From Fig. 27, we know that this function

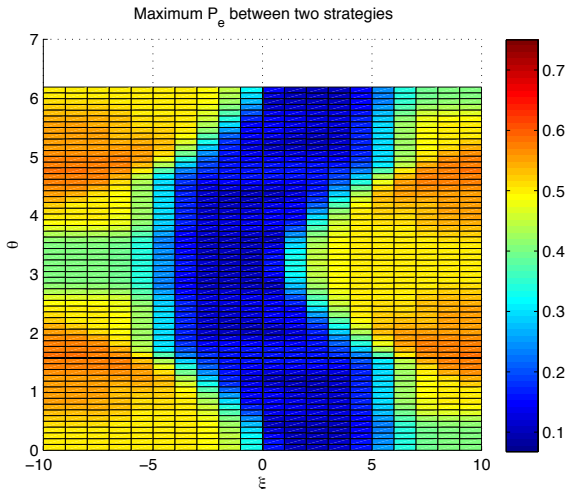

Fig. 25. Maximum P_e between two strategies shown in ξ-θ space when $\eta = 0$, where $\frac{E_s}{E_j} = 0.5$, $\frac{E_s}{N_0} = 20$, and $N = 5$.

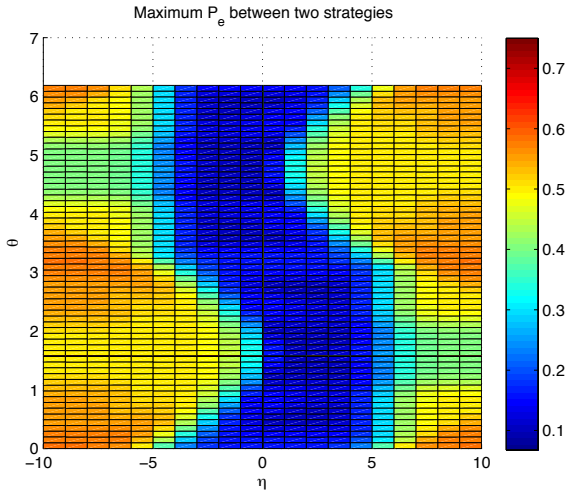

Fig. 26. Maximum P_e between two strategies shown in η-θ space when $\xi = 0$, where $\frac{E_s}{E_j} = 0.5$, $\frac{E_s}{N_0} = 20$, and $N = 5$.

has minimum value when (ξ, η) is close to $(0, 0)$. This is why the optimum solution of (5.16) is $\xi = \eta = 0$ and $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$.

To compare the optimal value of P_e between two strategies and evaluate the performance under different settings, we change $\frac{E_s}{E_j}$, $\frac{E_s}{N_0}$, or N to get numerical results.

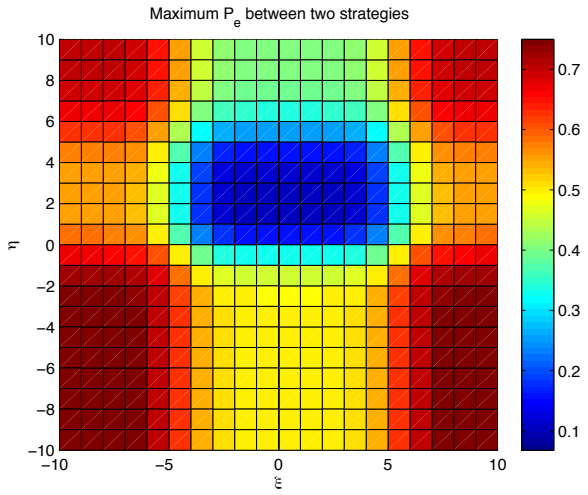

Fig. 27. Maximum P_e between two strategies shown in ξ-η space when $\theta = \frac{\pi}{2}$, where $\frac{E_s}{E_j} = 0.5$, $\frac{E_s}{N_0} = 20$, and $N = 5$.

By changing $\frac{E_s}{E_j}$, we have Fig. 28. The optimal value of P_e is inversely proportional to $\frac{E_s}{E_j}$ under both strategies. Furthermore, as $\frac{E_s}{E_j}$ increases, Strategy I becomes better choice for the jammer than Strategy II. In other words, it is better for the jammer to choose single-tone jamming when jamming power is low. We also know that anti-jammer can reduce symbol error probability by increasing signal power. By changing $\frac{E_s}{N_0}$, we get Fig. 29. The behavior of Strategy II under different $\frac{E_s}{N_0}$ is similar to that under different $\frac{E_s}{E_j}$. But the performance of Strategy I is nearly constant when SNR is between 8dB and 20dB. From Fig. 29, we also know that Strategy I is a better choice for the jammer when SNR is high and Strategy II is a better choice for the jammer when SNR is low. In this chapter, we define signal to interference plus noise ratio (SINR) as $\frac{E_s}{E_j+N_0}$. Figs. 30 and 31 show the optimal value of P_e vs. SINR, which are under the same settings as in Figs. 28 and 29 respectively.

By changing N, we obtain Fig. 32. The optimal value of P_e is inversely proportional to N under both strategies, which indicates that the anti-jammer is able to reduce symbol error probability by increasing the number of frequency hopping

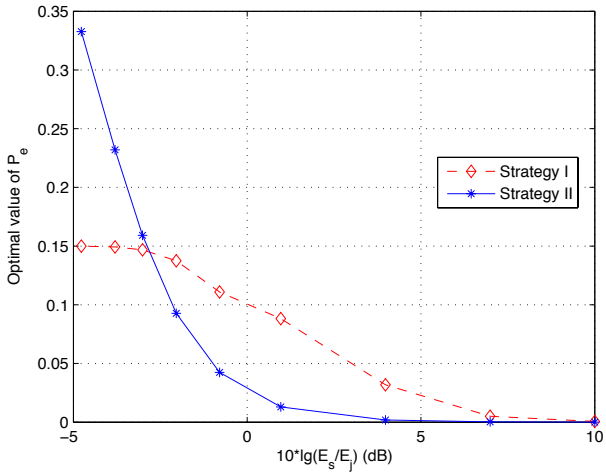

Fig. 28. Optimal value of P_e vs. $\frac{E_s}{E_j}$, where $\frac{E_s}{N_0} = 20$ and $N = 5$.

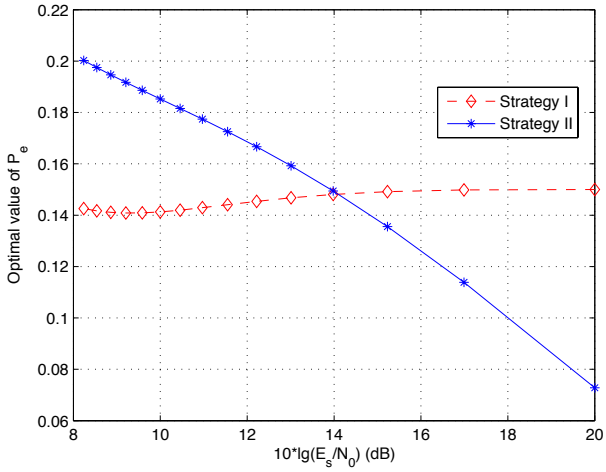

Fig. 29. Optimal value of P_e vs. $\frac{E_s}{N_0}$, where $\frac{E_s}{E_j} = 0.5$ and $N = 5$.

channels. From Fig. 32, we also know that Strategy I is preferred by the jammer when N is large and Strategy II is preferred by the jammer when N is small.

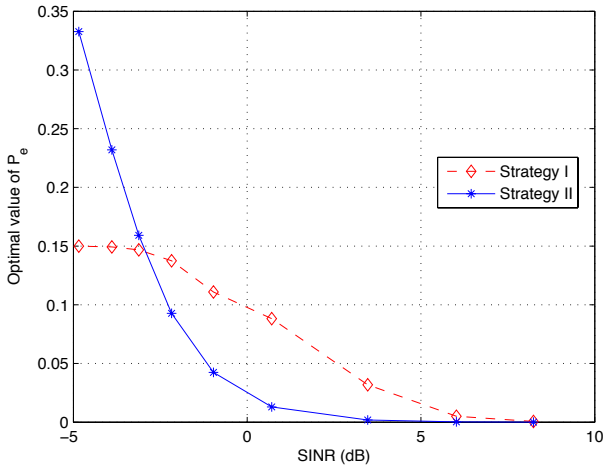

Fig. 30. Optimal value of P_e vs. SINR, where $\frac{E_s}{N_0} = 20$ and $N = 5$.

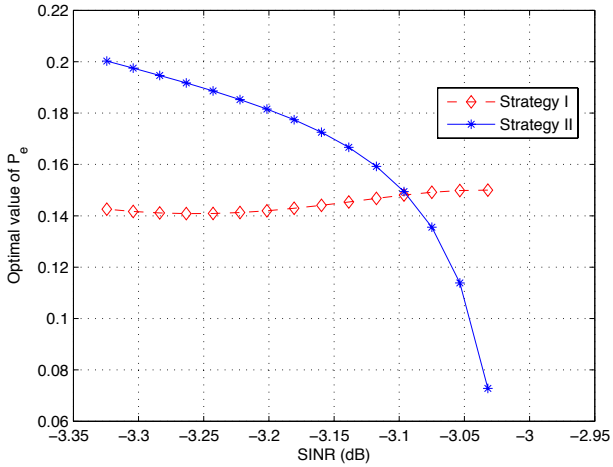

Fig. 31. Optimal value of P_e vs. SINR, where $\frac{E_s}{E_j} = 0.5$ and $N = 5$.

5.5 Conclusion

In this chapter, we designed a minimax anti-jammer for FHSS/QPSK satellite communication system. Two jamming strategies are evaluated: single-tone jamming and full-band multitone jamming. For each strategy, the phase of the jamming signal can be changed. The anti-jammer resists attacks from the jammer by choosing the

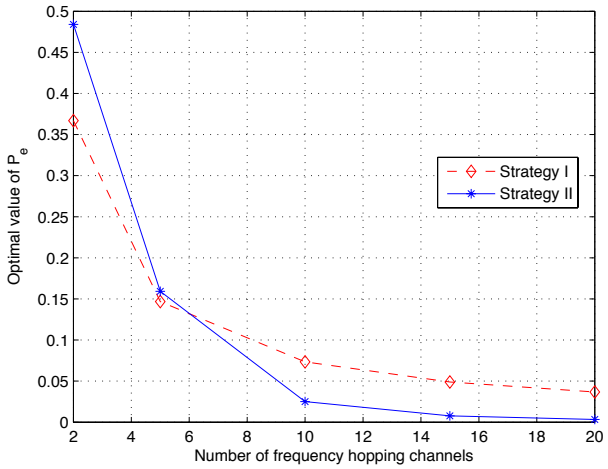

Fig. 32. Optimal value of P_e vs. number of frequency hopping channels, where $\frac{E_s}{E_j} = 0.5$ and $\frac{E_s}{N_0} = 20$.

optimal decision thresholds. The symbol error probability corresponding to each jamming strategy with different jamming phases when different decision thresholds are used is derived. By solving this minimax optimization problem, we know that the best decision thresholds for the anti-jammer when it has no prior knowledge of the phase of the jamming signal are the same as those when received signals are not attacked by the jammer, and the anti-jammer can reduce the symbol error probability by increasing signal power or employing more frequency hopping channels. If the jammer knows the anti-jammer's minimax decision thresholds, the best choice for the jammer is using single-tone jamming when the jamming power is low or N is large, and single-tone jamming will cause nearly constant P_e to the legitimate communication system under different SNRs.

CHAPTER 6

CONCLUSION

In this dissertation, joint detection-state estimation and secure signal processing were studied.

A new joint sequential object detection and system state estimation approach based on Wald's SPRT was proposed to achieve improved detection and system state estimation performance. This approach is weakly admissible and the probabilities of error can be controlled by the thresholds. The closed-form formulas for the first and second moments of the test statistic under both hypotheses were derived. Based that, both Kullback-Leibler distance and deflection coefficient were obtained. Numerical results showed that both Kullback-Leibler distance and deflection coefficient are monotonically increasing function of the number of samples which implies that both probability of termination and performance of the proposed algorithm are increasing with time. Numerical results also showed that the sequential detection algorithm detects a moving object with a small ASN and low probabilities of error even under low SNR conditions, and it outperforms the optimal FSS detector significantly in terms of the number of samples.

To guarantee that this approach will eventually terminate with probability one, a terminative joint sequential object detection and system state estimation approach based on a fused test statistic was proposed. This is a fused algorithm by using both dependent measurements and independent samples. The relationship between fused algorithm and the Wald's SPRT with only one test statistic was derived in terms of probabilities of error. The upper bounds on the actual probabilities of error

were derived, which are represented via nominal probabilities of error. Therefore, the probabilities of error of this fused algorithm can also be controlled by the thresholds.

Besides detection and system state estimation, secure signal processing is also an important problem in surveillance systems. One interesting topic is how to reliably estimate system state when faulty sensors exist in large sensor networks especially when measurement noise is correlated. In this dissertation, joint group testing of time-varying faulty sensors and system state estimation in the presence of correlated measurement noise was proposed. Compared to the one-by-one test, the proposed approach has significantly better system state estimation performance and improves faulty sensor detection performance, especially the probability of miss. Furthermore, the proposed approach achieves almost the same state estimation performance as a clairvoyant Kalman filter with the perfect knowledge of the sensor fault state.

The secure signal processing is also important in satellite communication systems. To improve the system's robustness, a minimax anti-jammer was design for FHSS/QPSK satellite communication systems. The anti-jammer resists attacks from the jammer by choosing the optimal decision thresholds. The best decision thresholds for the anti-jammer when it has no prior knowledge of the phase of the jamming signal was evaluated and the strategies for anti-jammer to reduce the symbol error probability were provided.

REFERENCES

[1] A. Alagil, M. Alotaibi, and Y. Liu. "Randomized positioning DSSS for anti-jamming wireless communications". In: *2016 International Conference on Computing, Networking and Communications (ICNC)*. 2016, pp. 1–6.

[2] S. Fang, Y. Liu, and P. Ning. "Wireless Communications under Broadband Reactive Jamming Attacks". In: *IEEE Transactions on Dependable and Secure Computing* 13.3 (2016), pp. 394–408.

[3] S.C. Chan, H.C. Wu, and K.M. Tsui. "Robust Recursive Eigendecomposition and Subspace-Based Algorithms With Application to Fault Detection in Wireless Sensor Networks". In: *IEEE Transactions on Instrumentation and Measurement* 61.6 (2012), pp. 1703–1718.

[4] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. New York: Wiley, 2001.

[5] R.E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME-Journal of Basic Engineering* 82 (1 1960), pp. 35–45.

[6] S.J. Julier and J.K. Uhlmann. "A New Extension of the Kalman Filter to Nonlinear Systems". In: *Proc. of the AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*. Vol. 3068. Orlando, FL, 1997, 182193.

[7] M.S. Arulampalam et al. "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking". In: *IEEE Transactions on Signal Processing* 50.2 (2002), pp. 174–188.

[8] F. Daum and J. Huang. "Particle flow for nonlinear filters, Bayesian decisions and transport". In: *Proc. 16th International Conference on Information Fusion (FUSION)*. Istanbul, Turkey, 2013, pp. 1072–1079.

[9] S.M. Kay. *Fundamentals of Statistical Signal Processing: Detection theory*. Ed. by A.V. Oppenheim. reprint. New York: Prentice Hall PTR, 1998.

[10] G. Casella and R.L. Berger. *Statistical Inference*. 2nd. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002.

[11] L. Weiss. "Testing one Simple Hypothesis Against Another". In: *Ann. Math. Statist.* 24.2 (1953), pp. 273–281.

[12] H. Chernoff. *Sequential Analysis and Optimal Design*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1987.

[13] A. Wald. *Sequential Analysis*. New York: John Wiley and Sons, Inc., 1966.

[14] A. Wald and J. Wolfowitz. "Optimum Character of the Sequential Probability Ratio Rest". In: *Ann. Math. Statist.* 19.3 (1948), pp. 326–339.

[15] B. Eisenberg, B.K. Ghosh, and G. Simons. "Properties of Generalized Sequential Probability Ratio Tests". In: *The Annals of Statistics* 4.2 (1976), pp. 237–251.

[16] J. Cochlar and I. Vrana. "On the Optimum Sequential Test of Two Hypotheses for Statistically Dependent Observations". In: *Kybernetika* 14.1 (1978), pp. 57–69.

[17] J. Kiefer and L. Weiss. "Some Properties of Generalized Sequential Probability Ratio Tests". In: *Ann. Math. Statist.* 28.1 (1957), pp. 57–74.

[18] D. Middleton and R. Esposito. “Simultaneous Optimum Detection and Estimation of Signals in Noise”. In: *IEEE Transactions on Information Theory* 14.3 (1968), pp. 434–444.

[19] A. Fredriksen, D. Middleton, and D. VandeLinde. “Simultaneous Signal Detection and Estimation under Multiple Hypotheses”. In: *IEEE Transactions on Information Theory* 18.5 (1972), pp. 607–614.

[20] B. Baygun and A.O. Hero. “Optimal simultaneous detection and estimation under a false alarm constraint”. In: *IEEE Transactions on Information Theory* 41.3 (1995), pp. 688–703.

[21] G.V. Moustakides et al. “Joint Detection and Estimation: Optimum Tests and Applications”. In: *IEEE Transactions on Information Theory* 58.7 (2012), pp. 4215–4229.

[22] G.H. Jajamovich, A. Tajer, and X. Wang. “Minimax-Optimal Hypothesis Testing With Estimation-Dependent Costs”. In: *IEEE Transactions on Signal Processing* 60.12 (2012), pp. 6151–6165.

[23] S.D. Blostein and H.S. Richardson. “A Sequential Detection Approach to Target Tracking”. In: *IEEE Transactions on Aerospace and Electronic Systems* 30.1 (1994), pp. 197–212.

[24] S.D. Blostein and T.S. Huang. “Detecting Small, Moving Objects in Image Sequences Using Sequential Hypothesis Testing”. In: *IEEE Transactions on Signal Processing* 39.7 (1991), pp. 1611–1629.

[25] S.M. Tonissen and R.J. Evans. “Peformance of dynamic programming techniques for Track-Before-Detect”. In: *IEEE Transactions on Aerospace and Electronic Systems* 32.4 (1996), pp. 1440–1451.

[26] E. Grossi and M. Lops. "Sequential Along-Track Integration for Early Detection of Moving Targets". In: *IEEE Transactions on Signal Processing* 56.8 (2008), pp. 3969–3982.

[27] J. Yan et al. "Joint detection and tracking processing algorithm for target tracking in multiple radar system". In: *IEEE Sensors Journal* 15.11 (2015), pp. 6534–6541.

[28] Benoit Fortin, Régis Lherbier, and Jean-Charles Noyer. "A model-based joint detection and tracking approach for multi-vehicle tracking with lidar sensor". In: *IEEE Transactions on Intelligent Transportation Systems* 16.4 (2015), pp. 1883–1895.

[29] B.T. Vo et al. "Multi-Sensor Joint Detection and Tracking with the Bernoulli Filter". In: *IEEE Transactions on Aerospace and Electronic Systems* 48.2 (2012), pp. 1385–1402.

[30] R. Niu. "Joint Object Detection and Tracking in Sensor Networks". In: *Proc. 16th International Symposium on Wireless Personal Multimedia Communications (WPMC)*. Atlantic City, NJ, 2013, pp. 1–5.

[31] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. 2nd. A Wiley-Interscience publication. Wiley, 2006.

[32] V. Schmidt. *Stochastics III*. Tech. rep. Ulm University, 2013.

[33] R.C. Luo, R.E. Mullen Jr, and D.E. Wessell. "An adaptive robotic tracking system using optical flow". In: *Proc. IEEE International Conference on Robotics and Automation*. 1988, pp. 568–573.

[34] K. Kawamura et al. "Intelligent robotic systems in service of the disabled". In: *IEEE Transactions on Rehabilitation Engineering* 3.1 (1995), pp. 14–21.

[35] M.H. Heitger et al. “Eye movement and visuomotor arm movement deficits following mild closed head injury”. In: *Brain* 127.3 (2004), pp. 575–590.

[36] M.H. Heitger et al. “Motor deficits and recovery during the first year following mild closed head injury”. In: *Brain Injury* 20.8 (2006), pp. 807–824.

[37] R. Niu and P.K. Varshney. “Sampling Schemes for Sequential Detection with Dependent Observations”. In: *IEEE Transactions on Signal Processing* 58.3 (2010), pp. 1469–1481.

[38] S. Andrilli and D. Hecker. *Elementary Linear Algebra*. 4th. Elsevier Science, 2010.

[39] N.J.A. Harvey et al. “Non-Adaptive Fault Diagnosis for All-Optical Networks via Combinatorial Group Testing on Graphs”. In: *Proc. 26th IEEE International Conference on Computer Communications*. 2007, pp. 697–705.

[40] Y. Zhou, G. Xu, and Q. Zhang. “Overview of fault detection and identification for non-linear dynamic systems”. In: *Proc. IEEE International Conference on Information and Automation (ICIA)*. Hailar, China, 2014, pp. 1040–1045.

[41] W. Xue, Y. Guo, and X. Zhang. “A Bank of Kalman Filters and a Robust Kalman Filter Applied in Fault Diagnosis of Aircraft Engine Sensor/Actuator”. In: *Proc. 2nd International Conference on Innovative Computing, Information and Control (ICICIC)*. Kumamoto, Japan, 2007, pp. 10–14.

[42] J. Lu and R. Niu. “Sparse attacking strategies in multi-sensor dynamic systems maximizing state estimation errors”. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016, pp. 3151–3155.

[43] R. Niu and J. Lu. "False information detection with minimum mean squared errors for Bayesian estimation". In: *Proc. 49th Annual Conference on Information Sciences and Systems (CISS)*. 2015, pp. 1–6.

[44] J. Lu and R. Niu. "A state estimation and malicious attack game in multi-sensor dynamic systems". In: *Proc. 18th International Conference on Information Fusion (Fusion)*. 2015, pp. 932–936.

[45] Y. Zhang, C. Szabo, and Q. Z. Sheng. "An Estimation Maximization Based Approach for Finding Reliable Sensors in Environmental Sensing". In: *Proc. IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*. 2015, pp. 190–197.

[46] Kevin Ni et al. "Sensor Network Data Fault Types". In: *ACM Trans. Sen. Netw.* 5.3 (2009), 25:1–25:29. ISSN: 1550-4859. DOI: 10.1145/1525856.1525863. URL: http://doi.acm.org/10.1145/1525856.1525863.

[47] R. Dorfman. "The Detection of Defective Members of Large Populations". In: *The Annals of Mathematical Statistics* 14.4 (1943), pp. 436–440.

[48] D.L. Donoho. "Compressed Sensing". In: *IEEE Transactions on Information Theory* 52.4 (2006), pp. 1289–1306.

[49] R.G. Baraniuk. "Compressive Sensing [Lecture Notes]". In: *IEEE Signal Processing Magazine* 24.4 (2007), pp. 118–121.

[50] E.J. Candes and M.B. Wakin. "An Introduction to Compressive Sensing". In: *IEEE Signal Processing Magazine* 25.2 (2008), pp. 21–30.

[51] C. Lo et al. "Efficient sensor fault detection using combinatorial group testing". In: *Proc. IEEE International Conference on Distributed computing in sensor systems (DCOSS)*. Cambridge, MA, 2013, pp. 199–206.

[52] T. Tosic, N. Thomos, and P. Frossard. “Distributed Sensor Failure Detection in Sensor Networks”. In: *Signal Processing* 93.2 (2013), pp. 399–410.

[53] J.S. Golan. *The Linear Algebra a Beginning Graduate Student Ought to Know*. 3rd. Springer Netherlands, 2012.

[54] D. Malioutov and M. Malyutov. “Boolean compressed sensing: LP relaxation for group testing”. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Kyoto, Japan, 2012, pp. 3305–3308.

[55] G.K. Atia and V. Saligrama. “Boolean Compressed Sensing and Noisy Group Testing”. In: *IEEE Transactions on Information Theory* 58.3 (2012), pp. 1880–1901.

[56] M. Aldridge, L. Baldassini, and O. Johnson. “Group Testing Algorithms: Bounds and Simulations”. In: *IEEE Transactions on Information Theory* 60.6 (2014), pp. 3671–3687.

[57] Andrs Seb. “On two random search problems”. In: *Journal of Statistical Planning and Inference* 11.1 (1985), pp. 23 –31.

[58] S. Liu et al. “Sensor Selection for Estimation with Correlated Measurement Noise”. In: *IEEE Transactions on Signal Processing* 64.13 (2016), pp. 3509–3522.

[59] E. B. Felstead and R. J. Keightley. “Performance of Multiple-Access Frequency-Hopped Systems in the Presence of Spurious Tones”. In: *IEEE Military Communications conference*. 2006, pp. 1–7.

[60] G. Wang et al. “Spread spectrum design for aeronautical communication system with radio frequency interference”. In: *IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*. 2015, 2F1–1–2F1–11.

[61] R. Mathis and R. Pawula. “A Spread Spectrum System with Frequency Hopping and Sequentially Balanced Modulation - Part II: Operation in Jamming and Multipath”. In: *IEEE Transactions on Communications* (1980).

[62] M. Simon and A. Polydoros. “Coherent Detection of Frequency-Hopped Quadrature Modulations in the Presence of Jamming - Part I: QPSK and QASK Modulations”. In: *IEEE Transactions on Communications* 29.11 (1981), pp. 1644–1660.

[63] Q. Chen et al. “Error performance of coded SFH/DPSK in tone interference and AWGN”. In: *IEE Proceedings I - Communications, Speech and Vision* 140.4 (1993), pp. 262–268.

[64] R. Ghareeb and A. Yongacoglu. “Performance analysis of frequency hopped/coherent MPSK in the presence of multitone jamming”. In: *IEEE Transactions on Communications* 44.2 (1996), pp. 152–155.

[65] Xin Tian et al. “Jamming/anti-jamming game with a cognitive jammer in space communication”. In: *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics. 2012, 83850Q–83850Q–10.

[66] J.G. Proakis and M. Salehi. *Fundamentals of Communication Systems*. 2nd. Pearson Education, 2013.

VITA

Mengqi Ren was born on May 18, 1987, in Xuchang, Henan Province, and is a Chinese citizen. She received her Bachelor of Engineering in Telecommunications and Bachelor of Science in Applied Psychology from East China Normal University (ECNU), Shanghai, China in 2009, and was awarded title of Excellent New College Graduate from ECNU. She received a Master of Science in Integrated Circuit and System from Peking University (PKU) in 2012, and was awarded title of Excellent Master Thesis from PKU.

Publications

Journal Papers

[1] M. Ren and Y. Zou, "A novel multiple sparse source localization using triangular pyramid microphone array," *IEEE Signal Processing Letters*, vol. 19, pp. 83-86, 2012.
[2] M. Ren, Y. Zou, H. Liu, and B. Li, "Binaural multiple sources localization with extracting high local SNR frequencies based on cepstrum," *International Proceedings of Computer Science and Information Technology*, vol. 51, pp. 729-734, November 2012.

Conference Papers

[1] M. Ren, R. Niu, G. Wang, and G. Chen, "Minimax anti-jammer design for FHSS/QPSK satellite communication systems," accepted by *Military Communications Conference*, Baltimore, MD, November 2016.
[2] M. Ren and R. Niu, "Terminative joint sequential object detection and tracking based on fused test statistics," *Proc. the 18th International Conference on Informa-*

tion Fusion, Washington, D.C., pp. 529-533, July 2015.

[3] P. Han, R. Niu, M. Ren, and Y. C. Eldar, "Distributed approximate message passing for sparse signal recovery," *Proc. the 2nd IEEE Global Conference on Signal and Information Processing: Information Processing for Big Data*, Atlanta, GA, pp. 497-501, December 2014.

[4] M. Ren and R. Niu, "A new joint sequential object detection and tracking approach and its performance analysis," *Proc. the 17th International Conference on Information Fusion*, Salamanca, Spain, pp. 1-7, July 2014.

[5] M. Ren and Y. Zou, "Performance analysis of multiple sparse source localization with triangular pyramid microphone array in noise environment," *Proc. IEEE International Conference on Signal Processing, Communication and Computing*, Hong Kong, pp. 315-319, August 2012.