Kalman Filtering with Intermittent Observations: Tail Distribution and Critical Value

Yilin Mo and Bruno Sinopoli

Abstract

In this paper we analyze the performance of Kalman filtering for linear Gaussian systems where packets containing observations are dropped according to a Markov process, modeling a Gilbert-Elliot channel. To address the challenges incurred by the loss of packets, we give a new definition of non-degeneracy, which is essentially stronger than the classical definition of observability but much weaker than one-step observability, which is usually used in the study of Kalman filtering with intermittent observations. We show that the trace of the Kalman estimation error covariance under intermittent observations follows a power decay law. Moreover we are able to compute the exact decay rate for non-degenerate systems. Finally we derive the critical value for non-degenerate systems based on the decay rate, which future improves the results from [1] and [2].

I. Introduction

A large wealth of applications demands wireless communication among small embedded devices. Wireless Sensor Network (WSN) technology provides the architectural paradigm to implement systems with a high degree of temporal and spatial granularity. Applications of sensor networks are becoming ubiquitous, ranging from environmental monitoring and control to building automation, surveillance and many others [3]. Given their low power nature and the requirement of long lasting deployment, communication between devices is limited in range and reliability. Changes in the environment, such as the simple relocation of a large metal object in a

This research is supported in part by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office. Foundation. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of ARO, CMU, or the U.S. Government or any of its agencies.

Yilin Mo and Bruno Sinopoli are with the ECE department of Carnegie Mellon University, Pittsburgh, PA ymo@andrew.cmu.edu, brunos@ece.cmu.edu

room or the presence of people, will inevitably affect the propagation properties of the wireless medium. Channels will be time-varying and unreliable. Spurred by this consideration, our effort concentrates on the design and analysis of estimation and control algorithms over unreliable networks.

A substantial body of literature has been devoted to such issues in the past few years. In this paper, we want to briefly revisit the paper of Sinopoli et al. [1]. In that paper, the authors analyze the problem of optimal state estimation for discrete-time linear Gaussian systems, under the assumption that observations are sent to the estimator via a memoryless erasure channel. This implies the existence of a non-unitary arrival probability associated with each packet. Consequently some observations will inevitably be lost. In this case although the discrete Kalman Filter [4] is still the optimal estimator, the boundedness of its error depends on the arrival probabilities of the observation packets. In particular the authors prove the existence of a critical arrival probability p_c, below which the expectation of estimation error covariance matrix P_k of Kalman filter will diverge. The authors are not able to compute the actual value of this critical probability for general linear systems, but provide upper and lower bounds. They are able to show that for special cases, for example when C matrix is invertible, the upper bound coincides with the lower bound and hence the exact critical value is obtained.

Philosophically such a phenomenon can be seen as related to the well known uncertainty principle [5], [6], which states that the optimal long-range control of a linear system with uncertain parameters does not exists if the uncertainty exceeds certain threshold. For Kalman filtering with intermittent observations, the uncertainty is incurred by the random packet loss and the optimal Kalman filter becomes unstable (i.e. the expectation of P_k is bounded) if too many packets are dropped.

Lots of research effort has been made to analyze the system with intermittent observations. One interesting direction is to characterize critical value for more general linear systems. Plarre and Bullo [2] relax the invertible condition on C matrix to C only invertible on the observable subspace. In [7], the authors prove that if the eigenvalues of system A matrix have distinguished absolute values, then the lower bound is indeed the critical value. The authors also provide a counter example to show that in general the lower bound is not tight.

The drawbacks of the above approach is that critical value only characterize the boundedness of the expectation. To completely characterize the impact of lossy network on state estimation, it is much more desirable to calculate the probability distribution of estimation error covariance matrix P_k instead of only considering the boundedness of its expectation. In [8], the author gives a closed-form expression for cumulative distribution function of P_k when the system satisfies non-overlapping conditions. In [9], the authors provide a numerical method to calculate the eigen-distribution of P_k under the assumption that the observation matrix C is random and time varying. In [10], the authors considered the probability of $P(P_k \leq M)$, and derived upper and lower bound on such probability. However, only in specific cases these upper and lower bound will coincides.

Other variations of the original problem are also considered. In [11], the authors introduce smart sensors, which send the local Kalman estimation instead of raw observation. In [12], a similar scenario is discussed where the sensor sends a linear combination of the current and previous measurement. A Markovian packet dropping model is introduced in [13] and a stability criterion is given. In [14], the authors study the case where the observation at each time splits into two parts, which are sent to the Kalman filter through two independent erasure channels. This work is further generalized by to n channels. A much more general model, which considered packet drop, delay and quantization of measurements at the same time, is introduced by Xie and Shi [15].

In the meantime, significant efforts have been made to design estimation and control schemes over lossy network, leveraging some of the results and methodologies mentioned above. Estimation of an unstable system is particularly important in control applications. Schenato et al. [16] show how an estimator rendered unstable by sensor data loss can render the closed loop system unstable. Similar packet drop models have been successfully used in sensor selection problems for estimation in sensor networks. In [17], the authors consider a stochastic sensor scheduling scheme, which randomly selected one sensor to transmit observation at each time. In [18], the authors show how to design the packet arrival rate to balance the state estimation error and energy cost of packet transmission.

This work breaks away from existing body of literature and uses a novel approach to characterize the stability of Kalman filtering under packet losses. There are several important contributions. First of all the problem formulation is general and can address any packet drop model. While existing results partially address the stability of the first moment of the error covariance under i.i.d packet losses, we characterize the stability of all moments under Markovian packet loss models.

We also show that the proposed methodology applies to general packet loss models, provided that certain probabilities associated with the specific loss process can be computed. This not only advances the current state of the art, as the computation of the critical value becomes a special case, but generalizes to a much larger class of erasure channel models. This goal is accomplished by characterizing the tail distribution of $trace(P_k)$, i.e. study how $P(trace(P_k) > M)$ converges to 0 as M goes to infinity. We will show that, under minimal assumptions on the system, $trace(P_k)$ follows a power decay law. This result has tremendous value of its own, as it provides the designer with a quantitative bound on the quality of the estimator. We are able compute the exact decay rate for a large class of linear Gaussian systems, defined as non-degenerate. We illustrate the relationship between non-degeneracy and observability and argue that such condition, rather than the weaker notion of observability, is the appropriate one to check when the observation process is random. Under this assumption we derive useful bounds for the Riccati equation, that can be used independently of any loss process. We then compute the critical value for non-degenerate system as a consequence of the tail distribution and prove that it attains the lower bound derived in [1]. Degenerate systems will in general have a higher critical probability, as shown in [7].

The paper is organized in the following manner: Section II formulates the problem. Section III defines non-degenerate system and compares it to the definition of observability and one-step observability. Section IV states several important inequalities on iterative Riccati and Lyapunov equations, which is used in Section V to derive the exact decay rate for non-degenerate systems. In Secion VI we derived the critical value and boundedness conditions of higher moments of P_k based on the tail distribution, and compare our results with the existing results from the literature. Finally Section VII concludes the paper.

II. Problem Formulation

Consider the following linear system

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \\ y_k &= Cx_k + v_k, \end{aligned} \tag{1}$$

where $x_k \in \mathbb{C}^n$ is the state vector, $y_k \in \mathbb{C}^m$ is the output vector, $w_k \in \mathbb{C}^n$ and $v_k \in \mathbb{C}^m$ are Gaussian random vectors[1] with zero mean and covariance matrices $Q \geq 0$ and $R > 0$, respectively. Assume that the initial state x_{-1} is also a Gaussian vector of mean $\bar{x}_{-1}$ and covariance matrix $\Sigma \geq 0$. Let w_i, v_i, x_{-1} be mutually independent. Define $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$, not necessarily distinct, as the eigenvalues of A.

Consider the case in which observations are sent to the estimator via a Gilbert-Elliot channel, where the packet arrival is modeled by a Markov process $\{\gamma_k\}$. According to this model, the measurement y_k sent at time k reaches its destination if $\gamma_k = 1$; it is lost otherwise. Let γ_k be independent of w_k, v_k, x_0, i.e. the communication channel is independent of both process and measurement noises and let the transition matrix to be

$$\begin{bmatrix} P(\gamma_{k+1}=0|\gamma_k=0) & P(\gamma_{k+1}=1|\gamma_k=0) \\ P(\gamma_{k+1}=0|\gamma_k=1) & P(\gamma_{k+1}=1|\gamma_k=1) \end{bmatrix} = \begin{bmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{bmatrix}.$$

Furthermore, we assume the Markov chain is irreducible and stationary, i.e. $0 < p_1 \leq 1$, $0 < p_2 \leq 1$ and $P(\gamma_0 = 0) = \ldots = P(\gamma_k = 0) = p_2/(p_1+p_2)$. In this paper we will also consider the i.i.d. packet drop model, which can be seen as a special case of Markovian model. It is easy to prove if $p_1 + p_2 = 1$, then γ_ks are i.i.d. distributed and we define $p = p_1$ in that case.

The Kalman Filter equations for this system were derived in [1] and take the following form:

$$\hat{x}_k = \hat{x}_{k|k-1} + \gamma_k K_k (y_k - C\hat{x}_{k|k-1}),$$
$$P_k = P_{k|k-1} - \gamma_k K_k C P_{k|k-1},$$

where

$$\hat{x}_{k+1|k} = A\hat{x}_k, \quad P_{k+1|k} = AP_kA^H + Q,$$
$$K_k = P_{k|k-1}C^H(CP_{k|k-1}C^H + R)^{-1},$$
$$\hat{x}_{-1} = \bar{x}_{-1}, \quad P_{-1} = \Sigma.$$

To simplify notations, let us first define the functions

$$h(X) \triangleq AXA^H + Q, \quad g(X) \triangleq h(X) - h(X)C^H(Ch(X)C^H + R)^{-1}Ch(X). \tag{2}$$

[1]Since we will use A in its diagonal standard form, we will assume every vector and matrix discussed in this paper to be in complex plane.

As a result, it is easy to see that

$$P_{k+1} = \begin{cases} h(P_k) & \gamma_{k+1} = 0 \\ g(P_k) & \gamma_{k+1} = 1 \end{cases} \tag{3}$$

It is easy to see that the following properties hold for both h and g:

Proposition 1: 1) $h(X_1) \geq h(X_2)$ if $X_1 \geq X_2$.

2) $g(X_1) \geq g(X_2)$ if $X_1 \geq X_2$.

3) $h(X) \geq g(X)$.

The proof can be found in the Proposition 3 in the appendix.

In this paper we wish to analyze the tail behavior of the distribution of P_k, i.e., we want to know how likely is for the Kalman filter to have a very large P_k due to packet loss. Let us first define

$$\varphi(M) \triangleq \sup_k P(trace(P_k) > M), \tag{4}$$

where $M > 0$ is a scalar. Hence, $\varphi(M)$ denotes the maximum probability of $trace(P_k)$ to be larger than M. However in general the exact value of $\varphi(M)$ is hard to compute. In this paper we are more concerned with the decay rate under which $\varphi(M)$ converges to 0 as M goes to infinity. Let us define the upper and lower decay rates respectively as

$$\overline{\phi} \triangleq \limsup_{M\to\infty} \frac{\log \varphi(M)}{\log M}, \ \underline{\phi} \triangleq \liminf_{M\to\infty} \frac{\log \varphi(M)}{\log M}. \tag{5}$$

It is easy to see that since $\varphi(M) \leq 1$, both $\overline{\phi}$ and $\underline{\phi}$ are always non-positive. If $\overline{\phi} = \underline{\phi}$, then we define $\phi \triangleq \overline{\phi} = \underline{\phi}$ as the decay rate.

Remark 1: If the decay rate $\phi > -\infty$ is well defined, we can conclude from the definition that the following inequality holds for sufficient large M:

$$M^{\phi-\delta} \leq \varphi(M) \leq M^{\phi+\delta},$$

where $\delta > 0$ can be arbitrarily small. As a result, we know the asymptotic behaviour when M approaches infinity. A slow decay rate would indicate that with high probability the filter will get a large estimation error covariance P_k over time, while a fast decay rate indicates that such event is less likely.

In this paper we also want to characterize the conditions under which $E(P_k^q)$ is uniformly bounded, where $q \in \mathbb{N}$. For the case where $q = 1$ and γ_ks are independent, Sinopoli et al. [1] prove the following existence theorem for the critical arrival probability:

Theorem 1: If $(A, Q^{\frac{1}{2}})$ *is controllable,* (C, A) *is detectable, and* A *is unstable, then there exists a critical value* $p_c \in [0, 1)$ *such that* [2,3]

$$\sup_k EP_k = +\infty \quad \textit{for } 0 \le p < p_c \ \textit{ and } \quad \textit{for some } P_0 \ge 0, \tag{6}$$

$$EP_k \le M_{P_0} \ \ \forall k \quad \textit{for } p_c < p \le 1 \ \textit{ and } \quad \textit{for all } P_0 \ge 0, \tag{7}$$

where $M_{P_0} > 0$ *depends on the initial condition* $P_0 \ge 0$*.*

For simplicity, we will say that EP_k is unbounded if $\sup_k EP_k = +\infty$ or EP_k is bounded if there exists a uniform bound independent of k. In the following sections, we will derive the decay rate of $\varphi(M)$ and use it to characterize the boundedness of $E(P_k^q)$.

III. Non-Degenerate Systems

In this section we will introduce the concept of non-degeneracy and provide some insight on why the new definition is crucial in the analysis of Kalman filtering with intermittent observations. We will also compare this new definition with the traditional definition of observability and one-step observability.

Before continuing on, we want to state the following assumptions on the system, which are assumed to hold throughout the rest of the paper:

$(H1)$ (C, A) is detectable.

$(H2)$ $(A,\, Q^{1/2})$ is controllable.

$(H3)$ A is unstable and diagonalizable[4].

Remark 2: The first assumption is essential for the classical Kalman filter to have a bounded estimation error. The second assumption is used to guarantee the existence of the critical value, as per Theorem 1, and is usually satisfied by the majority of real systems. We further require A to be unstable only because this is the only interesting case. If A is stable, one can prove that the Kalman filter will still have a bounded estimation error even in the absence of observations. The study of estimation of unstable systems is particularly important for closed loop control, as an unstable estimator can render a closed loop system unstable. Finally the requirement

[2]We use the notation $\sup_k A_k = +\infty$ when there is no matrix $M \ge 0$ such that $A_k \le M, \forall k$.

[3]Note that all the comparisons between matrices in this paper are in the sense of positive definite unless otherwise specified

[4]We allow complex eigenvalues.

of diagonalizability is used to define non-degenerate systems. While it is possible that such a requirement excludes some interesting systems, such as double integrators, we do believe most real systems are diagonalizable and therefore the results presented herein retian a great degree of generality. Although we believe that the results can be extended to Jordan forms, we expect the proofs to be lengthy, technical and not necessary insightful. Nonetheless, for the sake of completeness, we plan to remove this assumption and consider more general Jordan forms in future work.

Since we assume that A can be diagonalized, without loss of generality, we can assume it is already in the diagonal standard form by performing a linear transformation on the system. Also since the eigenvalues of A can be complex, we will use Hermitian instead of transpose in the rest of the paper.

We now are ready to define non-degenerate systems:

Definition 1: Consider the system (A, C) *in its diagonal standard form, i.e.* $A = diag(\lambda_1, \ldots, \lambda_n)$ *and* $C = [C_1, \ldots, C_n]$*. A block of the system is defined as a subsystem* $(A_{\mathcal{I}} = diag(\lambda_{i_1}, \ldots, \lambda_{i_j}), C_{\mathcal{I}} = [C_{i_1}, \ldots, C_{i_j}])$*,* $1 \leq i_1 < \ldots < i_j \leq n$*, where* $\mathcal{I} = \{i_1, \ldots, i_j\} \subset \{1, \ldots, n\}$ *is the index set.*

Definition 2: An equiblock is a block which satisfies $\lambda_{i_1} = \ldots = \lambda_{i_j}$*.*

Definition 3: A quasi-equiblock is a block which satisfies $|\lambda_{i_1}| = \ldots = |\lambda_{i_j}|$*.*

Definition 4: A system (A, C) *is one-step observable if* C *is full column rank.*

Definition 5: A diagonalizable system is non-degenerate if every quasi-equiblock of the system is one-step observable. It is degenerate if there exists at least one quasi-equiblock which is not one-step observable.

For example, if $A = diag(2, -2)$ and $C = [1, 1]$, then the system is degenerate since it is a quasi-equiblock and not one-step observable. For $A = diag(2, -2, 3, -3)$ and $C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$, the two equiblocks are $(diag(2, -2), I_2)$ and $(diag(3, -3), I_2)$ and both of them are one-step observable. Thus, the system is non-degenerate.

To compare observability and non-degeneracy, we provide the following theorem:

Theorem 2: A system is observable if and only if every equiblock is one-step observable.

Proof: This is a direct result of Theorem 1 in [19] and the duality between observability and controllability. ■

As a result, we can conclude that observability is weaker than non-degeneracy, since every equiblock is a quasi-equiblock. Therefore, a non-degenerate system is also observable. The converse is not true, as an observable system may be degenerate as shown in the example above).

On the other hand, non-degeneracy is obviously weaker than one-step observability. In fact, for a one-step observable system, the C matrix must have at least n independent rows, which implies y_k is at least a $\mathbb{C}^n$ vector, while, for a non-degenerate system, the C matrix is required only to have the number of rows equal to the dimension of the largest quasi-equiblock. Enforcing one-step observability in large systems requires the use of a potentially high number of sensors and communications at each sampling period. Enforcing non degeneracy in general requires the use of less sensors and communication bandwidth. We will show that the same asymptotic performance is attained for both one-step observable and non degenerate systems.

To summarize the above comparison, we have:

1) The system is observable if and only if every equiblock is one-step observable.
2) The system is non-degenerate if and only if every quasi-equiblock is one-step observable.
3) The system is one-step observable if C is full column rank.

Before proceeding, we wish to give some intuition on the importance of the concept of non-degeneracy for Kalman filtering with intermittent observations. The more rigorous discussion can be found in the next section. The main reason to introduce such concept is the loss of observability. It is well known that observability may be lost when discretizing a continuous time system or using a different sampling on a discrete time system, since different eigenvalues may rotate to the same point. The same thing happens when packet drops occur, which can be seen as a random sampling of the system. As a result, we need the stronger condition of non-degeneracy to ensure that no matter under what rate the system is sampled, it will always remain observable, which is illustrated by the following proposition:

Proposition 2: If the system is non-degenerate, then (A^q, C) is observable for any $q \in \mathbb{R}$.

Proof: We will use Theorem 2. First consider an equiblock $(A_{\mathcal{I}}^q, C_{\mathcal{I}})$, where

$$A_{\mathcal{I}}^q = diag(\lambda_{i_1}^q, \ldots, \lambda_{i_j}^q).$$

Since $\lambda_{i_1}^q = \ldots = \lambda_{i_j}^q$, we know that

$$|\lambda_{i_1}| = \ldots = |\lambda_{i_j}|.$$

As a result, $(A_{\mathcal{I}}, C_{\mathcal{I}})$ is a quasi-equiblock for the original system. Hence, it is one-step observable, which implies that $(A_{\mathcal{I}}^q, C_{\mathcal{I}})$ is also one-step observable. Therefore, all the equiblocks of (A^q, C) are one-step observable. Using Theorem 2 we conclude the proof. ∎

IV. General Inequalities on Iterative Riccati and Lyapunov Equations

Before deriving the decay rate, we will establish several inequalities on iterative Riccati and Lyapunov equations, which will be used in the next section. We want to emphasize that such inequalities are derived independently of the packet loss model and can therefore be used in a more general context.

To simplify notations, let us define

$$g^1(X) = g(X),\ g^i(X) = g(g^{i-1}(X)),\ h^1(X) = h(X),\ h^i(X) = h(h^{i-1}(X)).$$

where functions g, h are defined in (2). Moreover we will simply write the composition of g^i and h^j as

$$g^i h^j(X) = g^i(h^j(X)).$$

The first inequality provides a lower bound for $h^i(X)$:

Theorem 3: Suppose the system satisfies assumptions $(H1) - (H3)$*, then the following inequality holds*

$$trace(h^i(X)) \geq \underline{\alpha}|\lambda_1|^{2i},\ i \geq n, \tag{8}$$

where X is positive semidefinite and $\underline{\alpha} > 0$ is a constant independent of X and i.

Proof: First let us consider $h^n(X)$, from the definition of h, we can simply write it as

$$h^n(X) = Q + AQA^H + \cdots + A^{n-1}QA^{(n-1)H} + A^n X A^{nH},$$

where $A^{nH} = (A^H)^n$. Since we already assumed that $(A,\, Q^{1/2})$ is controllable, we know that $Q + \cdots + A^{n-1}QA^{(n-1)H}$ is strictly positive definite. Hence, it is possible to find $\underline{\alpha} > 0$, such that

$$Q + AQA^H + \cdots + A^{n-1}QA^{(n-1)H} > \underline{\alpha}|\lambda_1|^{2n} I_n > 0.$$

Now consider $h^i(X)$, where $i \geq n$,

$$h^i(X) = Q + AQA^H + \cdots + A^{i-1}QA^{(i-1)H} + A^i X A^{iH} \geq A^{i-n}QA^{(i-n)H} + \ldots A^{i-1}QA^{i-1}$$
$$= A^{i-n}\left(\sum_{j=0}^{n-1} A^j Q A^{jH}\right) A^{(i-n)H} \geq \underline{\alpha}|\lambda_1|^{2n} A^{i-n} A^{(i-n)H}.$$

Taking the trace on both sides and using the fact that A is a diagonal matrix, we have

$$trace(h^i(X)) \geq \underline{\alpha}|\lambda_1|^{2n}\left(\sum_{j=1}^{n}|\lambda_j|^{2i-2n}\right) \geq \underline{\alpha}|\lambda_1|^{2i},$$

which concludes the proof. ■

The following theorem characterizes the upper bound:

Theorem 4: Consider a system that satisfies assumptions $(H1)-(H3)$*. If the unstable part is non-degenerate, then the following inequality holds:*

$$trace(h^{i_1}gh^{i_2-1}g\cdots h^{i_l-1}gh^{i_{l+1}-1}(\Sigma)) \leq \overline{\alpha}\prod_{j=1}^{l}(|\lambda_j|+\varepsilon)^{2i_j}, \tag{9}$$

where l *is the number of unstable eigenvalues of the system,* $\overline{\alpha}>0$ *is a constant independent of* i_j *and* $\varepsilon>0$ *can be arbitrarily small.*

Proof: The proof is quite long and is reported in the appendix. ■

Remark 3: Suppose that l observations are received at times $k_1 > k_2 > \ldots > k_l$, where $k \geq k_1$ and $k_l \geq 0$. To simplify notations let us define $k_0 = k$ and $k_{l+1} = 0$ and $i_j = k_{j-1} - k_j$, $j = 1, 2, \ldots, l, l+1$. From the definition of h and g, we know that

$$P_k = h^{i_1}gh^{i_2-1}g\cdots h^{i_l-1}g(P_{k_l-1}),$$

and

$$P_{k_l-1} \leq h^{i_{l+1}-1}(\Sigma).$$

From Theorem 4, we know that the trace of P_k is bounded by $\overline{\alpha}\prod_{j=1}^{l}(|\lambda_j|+\varepsilon)^{2i_j}$. In other words, no matter how many packets are lost from time 1 to time k_l-1, we could always have a bounded estimation error as long as l packets arrive from time k_l to time k. In the classical setting of perfect communications, it is easy to derive from the observability Grammian that the estimation error is bounded when n sequential packets are received, provided the system is detectable. Hence, our result can be seen as a generalization of the classical result, since we do not require sequential packets arrivals. However, we do need to enforce the requirement of non-degeneracy, instead of observability, on the system.

To see why the requirement of non-degeneracy is crucial for (9) to hold, let us consider the system, where

$$A = \begin{bmatrix} 0 & \sqrt{2} \\ \sqrt{2} & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \end{bmatrix}, Q = \Sigma = I_2,\ R = 1.$$

It is easy to check the system satisfies assumptions $(H1)-(H3)$. However it is degenerate. Let $X = diag(a, b)$. Simple algebraic calculation yields

$$h(X) = diag(2b+1, 2a+1),\ g(X) = diag((2b+1)/(2b+2), 2a+1) \geq diag(0, 2a+1).$$

As a result, it is easy to see that

$$h^{i_1} g h^{i_2-1} g h^{i_3-1}(\Sigma) \geq diag(0, 2^{i_1+i_2+i_3+1}-1)$$

when both i_1 and i_2 are even. Hence inequality (9) does not hold since the right hand side depends not only on i_1 and i_2 but also i_3. In other word, we cannot bound the estimation error by a function of i_1 and i_2 even when two packets have arrived.

Also note that in the statement of the theorem we do not assume any channel failure models. As a result, the above theorem can also be used to analyze the performance of Kalman filtering with other packet drop models.

V. Decay rate for Non-degenerate Systems

In this section, we will use Theorem 3 and 4 to derive the decay rate for non-degenerate systems. Let us first define the following event:

$$E_{k,a}^{N} = \{\text{No more than } a \text{ packets arrive between time } k \text{ and time } k-N+1\},$$

where $k \geq N$.

Theorem 5: Suppose the system satisfies assumptions $(H1)-(H3)$*, then*

$$\underline{\phi} \geq \frac{\log(1-p_1)}{2\log|\lambda_1|}. \tag{10}$$

Proof: If $|\lambda_1| = 1$, then (10) becomes

$$\underline{\phi} \geq -\infty,$$

which will always hold. Hence, without loss of generality, let us assume $|\lambda_1| > 1$. First let us set M large enough and choose $N \in \mathbb{N}$ such that

$$\underline{\alpha}|\lambda_1|^{2N-2} \leq M < \underline{\alpha}|\lambda_1|^{2N}.$$

Since $|\lambda_1| > 1$, we can always find such N. Let us consider estimation error at time $k > N$. Suppose the observations made at time $k-N+1, \ldots, k$ are all lost, then by definition of h, we know that

$$P_k = h^N(P_{k-N}).$$

Now by Theorem 3, we know that

$$trace(P_k) \geq \underline{\alpha}|\lambda_1|^{2N} > M.$$

As a result, when packets $k-N+1$ to k are all lost, $trace(P_k) > M$. Hence

$$\varphi(M) = \sup_k P(trace(P_k) > M) \geq P(E_{k,0}^N)$$

The probability on the right hand side can be easily computed from the Markovian packet loss assumption:

$$P(E_{k,0}^N) = P(\gamma_{k-N+1} = \ldots = \gamma_k = 0) = P(\gamma_{k-N+1} = 0) \prod_{i=k-N+1}^{k-1} P(\gamma_{i+1} = 0|\gamma_i = 0)$$
$$= \frac{p_2}{p_1+p_2}(1-p_1)^{N-1}.$$

Since we already assumed that

$$\underline{\alpha}|\lambda_1|^{2N-2} \leq M,$$

we know that

$$N \leq \frac{\log M - \log \underline{\alpha}}{2\log|\lambda_1|} - 1.$$

As a result,

$$\underline{\phi} = \liminf_{M\to\infty} \frac{\log\varphi(M)}{\log M} \geq \liminf_{M\to\infty} \frac{\log(1-p_1)}{\log M}\left(\frac{\log M - \log\underline{\alpha}}{2\log|\lambda_1|} - 2\right) + \liminf_{M\to\infty} \frac{\log p_2 - \log(p_1+p_2)}{\log M}$$
$$= \frac{\log(1-p_1)}{2\log|\lambda_1|},$$

which concludes the proof. ■

Remark 4: The above theorem indicates that the distribution of $trace(P_k)$ follows a power decay rule as it can decay at most $M^{\underline{\phi}-\delta}$ fast when A is strictly unstable. Also since we do not assume non-degeneracy in the proof, the result is valid for general linear systems satisfying assumptions $(H1) - (H3)$.

Now let us consider an upper bound for $\overline{\phi}$, which is given by the following theorem:

Theorem 6: Consider a system which satisfies assumptions $(H1) - (H3)$ *and the unstable part is non-degenerate, then the following inequality holds:*

$$\overline{\phi} \leq \frac{\log(1-p_1)}{2\log|\lambda_1|}. \tag{11}$$

Proof: First we claim that

$$trace(P_k) \leq \beta|\lambda_1|^{2k},$$

where $\beta > 0$ is a constant independent of k. It is clear that the worst case happens when all the packets from time 1 to time k are lost. Hence

$$P_k \leq h^k(\Sigma) = \sum_{j=0}^{k-1} A^j Q A^{jH} + A^k \Sigma A^{kH}.$$

Now we can find a constant $\beta' > 0$, such that $Q \leq \beta' I_n$ and $\Sigma \leq \beta' I_n$. Therefore, we have

$$trace(P_k) \leq \beta' trace(\sum_{j=0}^{k} A^j A^{jH}) = \beta' \sum_{i=1}^{n} \frac{|\lambda_i|^{2k+2}-1}{|\lambda_i|^2-1} \leq n\beta' \frac{|\lambda_1|^{2k+2}-1}{|\lambda_1|^2-1} \leq \frac{n\beta'|\lambda_1|^2}{|\lambda_1|^2-1}|\lambda_1|^{2k},$$

where we use the fact that A is diagonal. Hence, $trace(P_k) \leq \beta|\lambda_1|^{2k}$, where $\beta = (n\beta'|\lambda_1|^2)/(|\lambda_1|^2 - 1)$.

Now let us fix $\varepsilon > 0$ defined in Theorem 4 and M large enough. Without loss of generality, we will also assume $\overline{\alpha} > \beta$, where $\overline{\alpha}$ is also defined in Theorem 4, since we could always pick a larger $\overline{\alpha}$. Choose N such that

$$\overline{\alpha}(|\lambda_1| + \varepsilon)^{2N} \leq M < \overline{\alpha}(|\lambda_1| + \varepsilon)^{2N+2}. \tag{12}$$

Since we assumed that $\overline{\alpha} > \beta$, we know that if $k \leq N$, then

$$trace(P_k) \leq \beta|\lambda_1|^{2k} \leq \beta|\lambda_1|^{2N} < M.$$

As a result, $P(trace(P_k) > M) = 0$, when $k \leq N$. Therefore

$$\varphi(M) = \sup_{k>N} P(trace(P_k) > M).$$

Now let us consider the case where $k > N$. Suppose we have received more than l packets from time $k - N + 1$ to time k. Assume that the latest packet is received at time k_1, and the second latest packet is received at time k_2, and so on. To simplify notation, define $k_0 = k$, $k_{l+1} = 0$ and $i_j = k_{j-1} - k_j$, $j = 1, \ldots, l+1$. From the definition of h and g, we know that

$$P_k = h^{i_1} g h^{i_2 - 1} g \cdots h^{i_l - 1} g(P_{k_l - 1}),$$

and

$$P_{k_l - 1} \leq h^{i_{l+1}-1}(\Sigma).$$

From Theorem 4, we know that

$$trace(P_k) = trace(h^{i_1} g h^{i_2-1} g \cdots h^{i_l-1} g h^{i_{l+1}-1}(\Sigma)) \leq \overline{\alpha} \prod_{j=1}^{l} (|\lambda_j| + \varepsilon)^{2i_j} \leq \overline{\alpha}(|\lambda_1| + \varepsilon)^{2(k-k_l)}.$$

Using the fact that more than l packets are received from time $k - N + 1$ to time k, we know that

$$k_l \geq k - N + 1 \Rightarrow k - k_l < N,$$

which implies that

$$trace(P_k) < \overline{\alpha}(|\lambda_1| + \varepsilon)^{2N} \leq M.$$

As a result, if more than l packets are received from time $k-N+1$ to time k, then $trace(P_k) \leq M$. Therefore

$$P(trace(P_k) > M) \leq P(E_{k,l-1}^N).$$

Now let us estimate the probability on the right hand side. Unlike the $P(E_{k,0}^N)$, the exact value of such probability cannot be easily computed. As a result we will only focus on the upper bound of $P(E_{k,l-1}^N)$. We know that there are totally 2^N possible combinations of $\gamma_{k-N+1}, \ldots, \gamma_k$ and $\sum_{i=0}^{l-1} \begin{pmatrix} N \\ i \end{pmatrix}$ of them belongs to the event $E_{k,l-1}^N$. Suppose that $\gamma'_{k-N+1}, \ldots, \gamma'_k$ is one of the eligible combination. As a result

$$P(\gamma'_{k-N+1}, \ldots, \gamma'_k) = P(\gamma'_{k-N+1}) \prod_{i=k-N}^{k} P(\gamma'_i | \gamma'_{i-1}).$$

Since there are at most $l - 1$ γ'_is which are 1, one can prove there exist at least $N - 2l + 1$ pairs of $(\gamma'_i, \gamma'_{i+1})$, which are both 0. As a result, the probability to get $(\gamma'_{k-N+1}, \ldots, \gamma'_k)$ is upper bounded by

$$P(\gamma'_{k-N+1}, \ldots, \gamma'_k) \leq (1 - p_1)^{N-2l+1}.$$

Therefore

$$P(E_{k,l-1}^N) \leq \sum_{i=0}^{l-1} \begin{pmatrix} N \\ i \end{pmatrix} (1 - p_1)^{-2l+1} (1 - p_1)^N = Poly(N)(1 - p_1)^N.$$

where

$$Poly(N) = \sum_{i=0}^{l-1} \begin{pmatrix} N \\ i \end{pmatrix} (1 - p_1)^{-2l+1}$$

is a $(l-1)$th polynomial of N. Therefore,

$$\overline{\phi} = \limsup_{M\to\infty} \frac{\log \varphi(M)}{\log M} \le \limsup_{M\to\infty} \frac{\log[Poly(N)]}{\log M} + \limsup_{M\to\infty} \frac{N\log(1-p_1)}{\log M}.$$

For the first term, by the first inequality in (12), we have

$$\limsup_{M\to\infty} \left|\frac{\log[Poly(N)]}{\log M}\right| \le \limsup_{N\to\infty} \left|\frac{\log[Poly(N)]}{2N\log(|\lambda_1|+\varepsilon)}\right| = 0.$$

For the second term, by the second inequality in (12), it is easy to establish the following inequality

$$N > \frac{\log M}{2\log(|\lambda_1|+\varepsilon)} - 1.$$

Therefore,

$$\limsup_{M\to\infty} \frac{N\log(1-p_1)}{\log M} \le \limsup_{M\to\infty} \left(\frac{\log M}{2\log(|\lambda_1|+\varepsilon)} - 1\right)\frac{\log(1-p_1)}{\log M} = \frac{\log(1-p_1)}{2\log(|\lambda_1|+\varepsilon)}.$$

Thus, we can conclude that

$$\overline{\phi} \le \frac{\log(1-p_1)}{2\log(|\lambda_1|+\varepsilon)}.$$

If we let ε goes to 0, we can conclude the proof. ■

Combining the above two Theorems, we have

Theorem 7: Consider a system which satisfies assumptions $(H1)-(H3)$ and the unstable part is non-degenerate, then the decay rate is given by the following equality:

$$\phi = \frac{\log(1-p_1)}{2\log|\lambda_1|}. \tag{13}$$

Remark 5: The results we derived here can be used for much more general packet drop models, as long as $P(E^n_{k,0})$ and $P(E^N_{k,l-1})$ can be computed.

Remark 6: For non-degenerate system, the decay rate is a decreasing function of p_1, which is very intuitive since larger p_1 indicates the channel is more capable of recovering from bad state. However this may not be true for degenerate system. For example, consider the system $A = diag(2,-2)$, $C = [1,1]$ and transition matrix

$$\left[\begin{array}{cc} P(\gamma_{k+1}=0|\gamma_k=0) & P(\gamma_{k+1}=1|\gamma_k=0) \\ P(\gamma_{k+1}=0|\gamma_k=1) & P(\gamma_{k+1}=1|\gamma_k=1) \end{array}\right] = \left[\begin{array}{cc} 1-p_1 & p_1 \\ 1 & 0 \end{array}\right].$$

If $p_1 = 1$, then the sequence of γ_ks will become $101010\cdots$. Hence, it is equivalent to sampling the system only at even (or odd depends on the γ_0) time and it is easy to see such system is unobservable. As a result $\varphi(M) = 1$ for any M and the decay rate is 0. However, if p_1 is not 1, then we can get several consecutive 0s in the sequence of γ_k. As a result, it is possible to get observations made both at odd and even time (for example it is possible to get "1001" for γ_0 to γ_3) and hence $\phi \leq 0$.

As discussed in the Section III, we know that the following inequality holds when M is sufficient large:

$$M^{\phi-\delta} \leq \sup_k P(trace(P_k) > M) \leq M^{\phi+\delta},$$

where $\delta > 0$ can be arbitrarily small. Such result is very useful to characterize the tail distribution of P_k. In the next section, we will apply this result to derive the boundedness conditions for higher moments of P_k and critical value for i.i.d. packet drop model.

VI. Critical Value and Boundedness of Higher Moments

In this section we want to derive the critical value and boundedness conditions of higher moments for Kalman filtering with intermittent observations based on the decay rate we derive in Section V, which is given by the following theorem:

Theorem 8: Consider a system which satisfies assumptions $(H1) - (H3)$ *and the unstable stable part is non-degenerate. Let* $q \in \mathbb{N}$. $E\left[P_k^q\right]$ *is uniformly bounded if*

$$p_1 > 1 - \frac{1}{|\lambda_1|^{2q}}. \tag{14}$$

It is unbounded for some initial condition Σ *if*

$$p_1 < 1 - \frac{1}{|\lambda_1|^{2q}}. \tag{15}$$

Proof: Let us first prove the case where $q = 1$. It is easy to see that EP_k is uniformly bounded if and only if $E[trace(P_k)]$ is uniformly bounded. As a result, we will only focus on the boundedness of $E[trace(P_k)]$. Now let us first prove that if $p_1 \geq 1 - |\lambda_1|^{-2}$, then $E[trace(P_k)]$ is uniformly bounded. Since $trace(P_k)$ is non-negative, it is easy to see the following equality holds

$$E\left[trace(P_k)\right] = \int_0^\infty P(trace(P_k) > M)dM.$$

Taking the supremum on both side yields

$$\sup_k E\left[trace(P_k)\right] = \int_0^\infty \varphi(M)dM.$$

Now by the definition of decay rate ϕ, we know that

$$\varphi(M) \le M^{\phi+\delta},$$

provided that M is sufficient large and $\delta > 0$ can be arbitrarily small. Hence, it is easy to check that the following condition is sufficient for $trace(EP_k)$ to be uniformly bounded:

$$\phi + \delta < -1,$$

which is equivalent to

$$\frac{\log(1-p_1)}{2\log|\lambda_1|} < -1 - \delta.$$

Manipulating the above equations we know the following condition is sufficient

$$p_1 > 1 - |\lambda_1|^{-2-2\delta}.$$

Since δ can be arbitrarily close to 0, we can conclude that

$$p_1 > 1 - |\lambda_1|^{-2},$$

is sufficient for $E(P_k)$ to be uniformly bounded. Then let us prove that $p_1 \le 1-|\lambda_1|^2$ is sufficient for an unbounded $E[trace(P_k)]$. Since $trace(P_k)$ is non-negative, it is easy to see that

$$E[trace(P_k)] \ge M \times P(trace(P_k) \ge M), \text{ for all } M.$$

Take the supremum over k and M on both side we have

$$\sup_k E[trace(P_k)] \ge \sup_M \left[M \times \varphi(M)\right]. \tag{16}$$

Since we know that

$$\varphi(M) \ge M^{\phi-\delta},$$

when M is sufficient large, the right hand side of (16) is unbounded if

$$1 + \phi - \delta > 0,$$

which is equivalent to

$$p_1 < 1 - |\lambda_1|^{-2+2\delta}.$$

Since δ can be arbitrary small, we can conclude that

$$p_1 < 1 - |\lambda_1|^2,$$

is sufficient for an unbounded $E(P_k)$.

Now let us consider arbitrary q. Suppose the eigenvalues of P_k are $\xi_1 \geq \xi_2 \geq \ldots \geq \xi_n$. Hence,

$$\sum_{i=1}^{n} \xi_i = trace(P_k), \sum_{i=1}^{n} \xi_i^q = trace(P_k^q).$$

Since P_k is positive semidefinite, ξ_i are non-negative. As a result, it is easy to prove that

$$trace(P_k^q) \leq trace(P_k)^q, \tag{17}$$

and the equality holds only when $\xi_1 = trace(P_k)$ and $\xi_2 = \ldots = \xi_n = 0$. Moreover,

$$trace(P_k^q) \geq n\left(\frac{trace(P_k)}{n}\right)^q, \tag{18}$$

and the equality holds only when $\xi_1 = \ldots = \xi_n = trace(P_k)/n$. Let us define

$$\varphi_q(M) \triangleq \sup_k P(trace(P_k^q) > M),$$

and similarly

$$\overline{\phi}_q \triangleq \limsup_{M\to\infty} \frac{\log \varphi_q(M)}{\log M}, \ \underline{\phi}_q \triangleq \liminf_{M\to\infty} \frac{\log \varphi_q(M)}{\log M}.$$

Moreover $\phi_q \triangleq \overline{\phi}_q$ if $\overline{\phi}_q = \underline{\phi}_q$. It is easy to see from (17) and (18) that

$$\phi_q = \phi/q.$$

Therefore, using the same argument as the proof for the case $q = 1$, we conclude the proof. ■

Since independent packet drop is a special case for the Markovian packet drop, we have the following corollary:

Corollary 1: Consider a system which satisfies assumptions $(H1) - (H3)$ *and the unstable stable part is non-degenerate. If* γ_k*s are i.i.d. distributed and* $P(\gamma_k = 1) = p$*, then the critical value of the system is given by*

$$p_c = 1 - |\lambda_1|^{-2} \tag{19}$$

Before finishing this section, we want to compare our result on critical value, with the results from [1] and [2]. Let us assume that $A = diag(\lambda_1, \ldots, \lambda_n)$ and $C = [C_1, \ldots, C_n]$. In [1], the authors could derive the exact critical value in two cases:

1) C is invertible.
2) A has only one unstable eigenvalue.

If C is invertible, then the system is one-step observable and thus non-degenerate. For the second case, suppose that λ_1 is the only unstable eigenvalue, it is easy to see that C_1 could not be a zero vector otherwise the system is not detectable. As a result, the unstable part, which is defined as the block $A_{\mathcal{I}} = \lambda_1$, $C_{\mathcal{I}} = C_1$, is non-degenerate by the definition.

In [2], the authors derive the critical value under the assumption that C is invertible on the observable subspace. Since the system is detectable, then all the unstable eigenvalues must be observable. Suppose that $\lambda_1, \dots, \lambda_l$ are unstable eigenvalues and $\lambda_{l+1}, \dots, \lambda_{l'}$ are stable and observable eigenvalues. Then matrix $[C_1, \dots, C_{l'}]$ must be full column rank since C is invertible on the observable space, which implies that $[C_1, \dots, C_l]$ must be also full column rank. Hence, the unstable part of the system is one-step observable and thus non-degenerate.

In conclusion, for diagonalizable systems, all the cases discussed in [1] and [2] are included in Theorem 8.

Remark 7: Note that the critical value for degenerate system is not $1 - |\lambda_1|^{-2}$ in general, as in shown in [7]. This is caused by the loss of observability incured by packet drop. As a result, some packets contain only redundant information which does not improve the estimation on some mode of the system and in general more packets are needed to obtain a bounded estimation.

VII. Conclusions and Future Work

In this paper we address the problem of state estimation for a discrete-time linear Gaussian system where observations are communicated to the estimator via a memoryless erasure channel. We were able to characterize the tail distribution of the estimation error covariance for Markovian packet losses and compute the value of the critical probability for each of its moments. We introduced the concept of non-degeneracy and show how such concept, rather than the weaker notion of observability, is the appropriate property to check when observations are sampled according to a stochastic process. This analysis leaves open the analysis for degenerate and non diagonalizable systems. Although some important systems fall within this class, we argue that our analysis covers most systems and can be extended, with considerable effort, to systems that admit non diagonal Jordan forms. As for degenerate systems, some preliminary results are contained in [7]. We plan to complete the characterization of the problem by generalizing our approach in both directions.

VIII. ACKNOWLEDGMENTS

The authors gratefully acknowledge Professors Francesco Bullo and P. R. Kumar, Craig Robinson for the numerous interesting discussions on the topic.

REFERENCES

[1] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, “Kalman filtering with intermittent observations,” *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, September 2004.

[2] K. Plarre and F. Bullo, “On kalman filtering for detectable systems with intermittent observations,” *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 386–390, 2009.

[3] N. P. Mahalik, *Sensor Networks and Configuration*. Springer, 2007.

[4] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[5] M. Athans, R. Ku, and S. B. Gershwin, “The uncertainty threshold principle, some fundamental limitations of optimal decision making under dynamic uncertainty,” *IEEE Transactions on Automatic Control*, vol. 22, no. 3, pp. 491–495, June 1977.

[6] R. Ku and M. Athans, “Further results on the uncertainty threshold principle,” *IEEE Transactions on Automatic Control*, vol. 22(5), pp. 491–495, October 1977.

[7] Y. Mo and B. Sinopoli, “A characterization of the critical value for kalman filtering with intermittent observations,” in *Proceedings of IEEE Conference on Decision and Control.*, 2008.

[8] A. Censi, “On the performance of Kalman filtering with intermittent observations: a geometric approach with fractals,” in *Proceedings of the American Control Conference (ACC)*, 2009.

[9] A. Vakili and B. Hassibi, “On the steady-state performance of kalman filtering with intermittent observations for stable systems,” in *48th IEEE Conference on Decision and Control (CDC)*, Dec 2009.

[10] L. shi, M. Epstein, and R. Murray, “Kalman filtering over a packet-dropping network: a probabilistic perspective,” *IEEE Transactions on Automatic Control*, vol. 55, no. 3, pp. 594–604, March 2010.

[11] Y. Xu and J. Hespanha, “Estimation under controlled and uncontrolled communications in networked control systems,” in *Proceedings of the CDC-ECC*, Sevilla, Spain, December 2005, pp. 842–847.

[12] C. Robinson and P. R. Kumar, “Sending the most recent observation is not optimal in networked control: Linear temporal coding and towards the design of a control specific transport protocol,” in *Proc. of IEEE CDC*, Dec 2007, pp. 334–339.

[13] M. Huang and S. Dey, “Stability of kalman filtering with markovian packet losses,” *Automatica*, vol. 43, no. 4, pp. 598–607, 2007.

[14] X. Liu and A. Goldsmith, “Kalman filtering with partial observation losses,” in *Proceedings of IEEE Conference on Decision and Control*, vol. 4, Bahamas, December 2004, pp. 4180–4186.

[15] X. Lihua and L. Shi, “State estimation over unreliable network,” in *Asian Control Conference*, August 2009.

[16] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, “Foundations of control and estimation over lossy networks,” *Proc. of the IEEE*, vol. 95, no. 1, pp. 163–187, January 2007.

[17] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, “On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage,” *Automatica*, vol. 42, no. 2, pp. 251–260, 2006.

[18] L. Shi and L. Qiu, "State estimation over a network: Packet-dropping analysis and design," in *The 7th IEEE International Conference on Control and Automation*, Dec 2009.

[19] C. Chen and C. Desoer, "A proof of controllability of jordan form state equations," *IEEE Transactions on Automatic Control*, vol. 13, no. 2, pp. 195–196, Apr 1968.

[20] R. S. Varga, *Geršgorin and His Circles*. New York: Springer, 2004.

IX. APPENDIX

This section is devoted to proving Theorem 4. Before continue on, we would like to state the following properties of function g and h, which will be used in the proof:

Proposition 3: Let

$$h(X,A,Q) = AXA^H+Q,\ \tilde{g}(X,R) = X-XC^H(CXC^H+R)^{-1}CX,\ g(X,A,Q,R) = \tilde{g}(h(X,A,Q),R).$$

Then the following propositions hold:

1) h *is an increasing function of* X *and* Q[5].
2) g *is an increasing function of* X, Q, R.
3) $h(X) \geq g(X)$.
4) $h(X,\alpha A,Q) \geq h(X,A,Q)$ *when* $\alpha > 1$.
5) $g(X,\alpha A,Q,R) \geq g(X,A,Q,R)$ *when* $\alpha > 1$.

Proof:

1) The first proposition is trivial.
2) It is easy to see that $\tilde{g}$ is an increasing function of R and hence that g is also increasing of R. To prove that g is an increasing function of X, let us first use the matrix inversion lemma to write $\tilde{g}$ as

$$\tilde{g}(X,R) = (X^{-1}+CR^{-1}C^H)^{-1}.$$

It is easy to see that $\tilde{g}$ is increasing in X. Combining this with the fact that h is also an increasing function of X, we can conclude that g is increasing in X.

3)

$$h(X)-g(X) = h(X)C^H(Ch(X)C^H+R)^{-1}Ch(X) \geq 0.$$

4) It is easy to see that

$$h(X,\alpha A,Q)-h(X,A,Q) = (\alpha^2-1)AXA^H \geq 0.$$

[5] By increasing we mean if $X_1 \geq X_2$, i.e. X_1-X_2 is positive semidefinite, then $h(X_1)-h(X_2)$ is also positive semidefinite.

5) The proposition is true because of the previous proposition and the fact that g is increasing in X.

■

Now we want to use the above propositions to simplify the problem. First since g, h are increasing in Q, R and X, we could find $\alpha_1 > 0$ such that $Q \le \alpha_1 I_n$, $\Sigma \le \alpha_1 I_n$ and $R \le \alpha_1 I_m$. Since we are only concerned with the upper bound, without loss of generality we can replace Q, Σ and R by $\alpha_1 I_n$, $\alpha_1 I_n$ and $\alpha_1 I_m$.

Furthermore without loss of generality we can assume that there is no eigenvalue of A that is on the unit circle by replacing A by $(1+\delta)A$, where $\delta > 0$ can be arbitrary small due to the fact that

$$h(X, \alpha A, Q) \ge h(X, A, Q),\ g(X, \alpha A, Q, R) \ge g(X, A, Q, R), \quad \text{if } \alpha > 1.$$

Combining with the assumption that A can be diagonalized, we will assume that $A = diag(A_1, A_2)$, where A_i is diagonal and A_1 and A_2 contain the strictly unstable and stable eigenvalues of A respectively. We will also assume $C = [C_1, C_2]$, where C_is are of proper dimensions.

From definition of h and g, we know that

$$P_k = h^{i_1} g h^{i_2 - 1} g \cdots h^{i_l - 1} g h^{i_{l+1} - 1}(\Sigma),$$

provided that l packets arrive at time $k_1 > \ldots > k_l$. Also define $k_0 = k$, $k_{l+1} = 0$ and $i_j = k_{j-1} - k_j$ for $1 \le j \le l+1$. As a result, proving Theorem 4 is equivalent to showing that

$$trace(P_k) \le \overline{\alpha} \prod_{j=1}^{n} (|\lambda_j| + \varepsilon)^{2 i_j}.$$

To prove the above inequality, we will exploit the optimality of Kalman filtering. We will construct a linear filter whose error covariance satisfies (9). Therefore P_k must also satisfy (9) due to the optimality of Kalman filtering. To construct such an estimator, let us rewrite the system equations as

$$x_{k+1,1} = A_1 x_{k,1} + w_{k,1}$$

$$x_{k+1,2} = A_2 x_{k,2} + w_{k,2},$$

$$y_k = C_1 x_{k,1} + v_k + C_2 x_{k,2}$$

Since A_2 is stable, we can just use $\hat{x}_{k,2} = A_2^k \overline{x}_{0,2}$ as an unbiased estimate of $x_{k,2}$. The new system equations become

$$\begin{aligned} x_{k+1,1} &= A_1 x_{k,1} + w_{k,1} \\ y'_k &= (y_k - C_2 A_2^k \overline{x}_{0,2}) = C_1 x_{k,1} + v_k + C_2(x_{k,2} - A_2^k \overline{x}_{0,2}) \end{aligned}$$

To obtain an estimator for $x_{k,1}$, let us first write down the relation between y'_k and $x_{k,1}$ as:

$$\begin{bmatrix} \gamma_k y'_k \\ \vdots \\ \gamma_1 y'_1 \\ \bar{x}_{0,1} \end{bmatrix} = \begin{bmatrix} \gamma_k C_1 \\ \vdots \\ \gamma_1 C_1 A_1^{-k+1} \\ A_1^{-k} \end{bmatrix} x_{k,1} + \begin{bmatrix} \gamma_k v_k \\ \vdots \\ \gamma_1 v_1 \\ \bar{x}_{0,1} - x_{0,1} \end{bmatrix} + \begin{bmatrix} \gamma_k C_2(x_{k,2} - A_2^k \overline{x}_{0,2}) \\ \vdots \\ \gamma_1 C_2(x_{1,2} - \overline{x}_{1,2}) \\ 0 \end{bmatrix} \tag{20}$$

$$- \begin{bmatrix} 0 & \cdots & 0 & 0 \\ \gamma_{k-1} C_1 A_1^{-1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_1 C_1 A_1^{-k+1} & \cdots & \gamma_1 C_1 A_1^{-1} & 0 \\ A_1^{-k} & \cdots & A_1^{-2} & A_1^{-1} \end{bmatrix} \begin{bmatrix} w_{k-1,1} \\ \vdots \\ w_{0,1} \end{bmatrix}.$$

where $\gamma_i = 1$ if and only if $i = k_j$, $j = 1, \dots, l$. To write (20) in a more compact form, let us define the following quantities:

$$F_k \triangleq \begin{bmatrix} A_1^{-1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ A_1^{-k+1} & \cdots & A_1^{-1} & 0 \\ A_1^{-k} & \cdots & A_1^{-2} & A_1^{-1} \end{bmatrix} \in \mathbb{C}^{lk \times lk}. \tag{21}$$

$$G_k \triangleq \begin{bmatrix} 0 & \cdots & 0 & 0 \\ C_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & 0 \\ 0 & \cdots & C_1 & 0 \\ 0 & \cdots & 0 & I_l \end{bmatrix} \in \mathbb{C}^{(mk+l) \times lk}. \tag{22}$$

$$e_k \triangleq -G_k F_k \begin{bmatrix} w_{k-1,1} \\ \vdots \\ w_0, 1 \end{bmatrix} + \begin{bmatrix} v_k \\ \vdots \\ v_1 \\ \bar{x}_{0,1} - x_{0,1} \end{bmatrix} + \begin{bmatrix} C_2(x_{k,2} - A_2^k \overline{x}_{0,2}) \\ \vdots \\ C_2(x_{1,2} - \overline{x}_{1,2}) \\ 0 \end{bmatrix} \in \mathbb{C}^{mk+l}. \tag{23}$$

$$T_k \triangleq \begin{bmatrix} C_1 \\ \vdots \\ C_1 A_1^{-k+1} \\ A_1^{-k} \end{bmatrix} \in \mathbb{C}^{(mk+l)\times l}, Y_k' = \begin{bmatrix} y_k' \\ \vdots \\ y_1' \\ \bar{x}_{0,1} \end{bmatrix} \in \mathbb{C}^{mk+l}. \tag{24}$$

Define Γ_k as removing the zero rows from $diag(\gamma_k I_m, \gamma_{k-1} I_m, \ldots, \gamma_1 I_m, I_l)$, where $\gamma_i = 1$ if and only if $i = k_j$, $j = 1, \ldots, l$. Thus Γ_k is a $ml + l$ by $mk + l$ matrix. Also define

$$\widetilde{Y}_k' \triangleq \Gamma_k Y_k,\ \widetilde{T}_k \triangleq \Gamma_k T_k,\ \widetilde{e}_k \triangleq \Gamma_k e_k.$$

Now we can rewrite (20) in a more compact form as

$$\widetilde{Y}_k' = \widetilde{T}_k x_k + \widetilde{e}_k. \tag{25}$$

Since we assumed, without loss of generality, that R, Q, Σ_0 are all diagonal matrices, it is easy to see that $x_{k,2}$ and $x_{k,1}$ are mutual independent. As a result, one can easily prove that the following estimator of x_k is unbiased

$$\hat{x}_{k,1} = (\widetilde{T}_k^H Cov(\widetilde{e}_k)^{-1} \widetilde{T}_k)^{-1} T_k^H Cov(\widetilde{e}_k)^{-1} \widetilde{Y}_k',\ \hat{x}_{k,2} = A_2^k \overline{x}_{0,2}, \tag{26}$$

with covariance

$$P_k' = \begin{bmatrix} P_{k,1}' & P_{k,off}' \\ (P_{k,off}')^H & P_{k,2}' \end{bmatrix}, \tag{27}$$

where

$$P_{k,1}' = (\widetilde{T}_k^H Cov(\widetilde{e}_k)^{-1} \widetilde{T}_k)^{-1},\ P_{k,2}' = Cov(x_{k,2}) = A_2^k \Sigma_0 A_2^{kH} + \sum_{i=0}^{k-1} A_2^i Q A_2^{iH}. \tag{28}$$

Since we are only concerned with the trace of P_k', it is easy to see that

$$trace(P_k') = trace(P_{k,1}') + trace(P_{k,2}'),$$

where $trace(P_{k,2}')$ is uniformly bounded regardless of k. Therefore, in order to prove (9) we will focus exclusively on $trace(P_{k,1}')$. Combining this argument with the optimality argument of Kalman filtering, we only need to prove that

$$trace(P_{k,1}') = (\widetilde{T}_k^H Cov(\widetilde{e}_k)^{-1} \widetilde{T}_k)^{-1} \le \overline{\alpha} \prod_{j=1}^{n} (|\lambda_j| + \varepsilon)^{2i_j}.$$

Now we claim that the following lemma is true:

Lemma 1: If a system satisfies assumptions $(H1)-(H3)$, then $P'_{k,1}$ is bounded by

$$P'_{k,1} \leq \alpha_2 (\sum_{j=1}^{n} (A^{k-k_j})^H C^H C A^{k-k_j} + A^{-kH}(A^{-k}))^{-1}, \tag{29}$$

where $\alpha_2 \in \mathbb{R}$ is constant independent of i_j.

Proof: We will bound $Cov(\widetilde{e}_k)$ by a diagonal matrix. Since we assume that w_k, v_k, x_0 are mutually independent, it is easy to prove that

$$Cov(e_k) = G_k F_k Cov\left(\begin{bmatrix} w_{k-1,1} \\ \vdots \\ w_{0,1} \end{bmatrix}\right) F_k^H G_k^H + Cov\left(\begin{bmatrix} v_k \\ \vdots \\ v_1 \\ \bar{x}_{0,1} - x_{0,1} \end{bmatrix}\right) + Cov\left(\begin{bmatrix} C_2 x_{k,2} \\ \vdots \\ C_2 x_{1,2} \\ 0 \end{bmatrix}\right).$$

Let us consider the first term. Notice that

$$F_k^{-1} = \begin{bmatrix} A_1 & & & \\ -I_l & \ddots & & \\ & \ddots & A_1 & \\ & & -I_l & A_1 \end{bmatrix}.$$

Therefore,

$$(F_k F_k^H)^{-1} = \begin{bmatrix} A_1^H A_1 + I & -A_1 & & \\ -A_1^H & \ddots & \ddots & \\ & \ddots & A_1^H A_1 + I & -A_1 \\ & & -A_1^H & A_1^H A_1 \end{bmatrix}.$$

By Gershgorin's Circle Theorem [20], we know that all the eigenvalues of $(F_k F_k^H)^{-1}$ are located inside one of the following circles:$|\zeta - |\lambda_i|^2 - 1| = |\lambda_i|, |\zeta - |\lambda_i|^2 - 1| = 2|\lambda_i|, |\zeta - |\lambda_i|^2| = |\lambda_i|$, where ζs are the eigenvalues of $(F_k F_k^H)^{-1}$.

Since $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_l| > 1$, for each eigenvalue of $(F_k F_k^H)^{-1}$, the following holds:

$$\zeta \geq min\{|\lambda_i|^2 + 1 - |\lambda_i|, |\lambda_i|^2 + 1 - 2|\lambda_i|, |\lambda_i|^2 - |\lambda_i|\}, \tag{30}$$

Thus, $0 < (|\lambda_l| - 1)^2 \leq \zeta$, which in turn gives

$$F_k F_k^H \leq \frac{1}{(|\lambda_l| - 1)^2} I_{lk}.$$

Since we assume that $Q = \alpha_1 I_n$, we can prove that

$$G_k F_k Cov\left(\begin{bmatrix} w_{k-1,1} \\ \vdots \\ w_0, 1 \end{bmatrix}\right) F_k^H G_k^H = \alpha_1 G_k F_k F_k^H G_k^H \leq \frac{\alpha_1}{(|\lambda_l| - 1)^2} G_k G_k^H.$$

From the definition,

$$\frac{\alpha_1}{(|\lambda_l| - 1)^2} G_k G_k^H = \frac{\alpha_1}{(|\lambda_l| - 1)^2} diag(0, C_1 C_1^H, \ldots, C_1 C_1^H, I_l),$$

which is uniformly bounded by $\alpha_3 I_{mk+l}$.

Now let us consider the second term, since $R = \alpha_1 I_m$, it is trivial to see that

$$Cov\left(\begin{bmatrix} v_k \\ \vdots \\ v_1 \\ \bar{x}_{0,1} - x_{0,1} \end{bmatrix}\right) = \alpha_1 I_{mk+l}.$$

Now consider the last term, let us write $x_{k,2}$ as

$$\begin{bmatrix} x_{k,2} \\ \vdots \\ x_{1,2} \\ x_{0,2} \end{bmatrix} = \begin{bmatrix} I_{n-l} & \cdots & A_2^{k-1} & A_2^k \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_{n-l} & A_2 \\ 0 & \cdots & 0 & I_{n-l} \end{bmatrix} \begin{bmatrix} w_{k-1,2} \\ \vdots \\ w_{0,2} \\ x_{0,2} \end{bmatrix}.$$

As a result,

$$Cov\left(\begin{bmatrix} x_{k,2} \\ \vdots \\ x_{1,2} \\ x_{0,2} \end{bmatrix}\right) = \alpha_1 \begin{bmatrix} I_{n-l} & \cdots & A_2^{k-1} & A_2^k \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_{n-l} & A_2 \\ 0 & \cdots & 0 & I_{n-l} \end{bmatrix} \begin{bmatrix} I_{n-l} & \cdots & A_2^{k-1} & A_2^k \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I_{n-l} & A_2 \\ 0 & \cdots & 0 & I_{n-l} \end{bmatrix}^H \leq \frac{\alpha_1}{(1 - |\lambda_{l+1}|)^2} I_{(n-l)(k+1)},$$

where the proof of the last inequality is similar to the proof of $F_k F_k^H \leq (|\lambda_l| - 1)^{-2} I_{lk}$ and is omitted. Therefore it is easy to see that

$$Cov\left(\begin{bmatrix} C_2 x_{k,2} \\ \vdots \\ C_2 x_{1,2} \\ 0 \end{bmatrix}\right) \leq \alpha_4 I_{mk+l},$$

where α_4 is a constant. As a result, we have proved that

$$Cov(e_k) \le \alpha_2 I_{mk+l},$$

where $\alpha_2 = \alpha_3 + \alpha_1 + \alpha_4$. Moreover

$$Cov(\widetilde{e}_k) = \Gamma_k Cov(e_k) \Gamma_k^H \le \alpha_2 I_{ml+l}.$$

The above bound is independent of $i_1, \ldots, i_{l+1}$, which proves

$$P'_{k,1} = (\widetilde{T}_k^H Cov(\widetilde{e}_k)^{-1} \widetilde{T}_k)^{-1} \le \alpha_2 (\sum_{j=1}^{l} (A^{k-k_j})^H C^H C A^{k-k_j} + A^{-kH} A^{-k})^{-1}.$$

■

We will manipulate $\sum_{j=1}^{n} (A^{-i_j})^H C^H C A^{-i_j}$ to prove the upper bound by using cofactors for matrix inversion. Before continue, we need the following lemmas.

Lemma 2: For a non-degenerate system, it is possible to find a set of row vectors $L_1, L_2, \ldots, L_l$, such that $L_i C = [l_{i,0}, \ldots, l_{i,l}]$, where $l_{i,i} = 1$ and $l_{i,a} = 0$ if $|\lambda_i| = |\lambda_a|$ and $i \neq a$.

Proof: It is simple to show that the lemma holds by using Gaussian Elimination for every quasi-equiblock. ■

Lemma 3: Consider that $|\lambda_1| \ge |\lambda_2| \cdots \ge |\lambda_l|$, $l_{i,i} = 1$ and $l_{i,a} = 0$ if $i \neq a$ and $|\lambda_i| = |\lambda_a|$. Let $i_1 = k - k_1$ and $i_j = k_{j-1} - k_j$ for $2 \le j \le l$. Define

$$D = \begin{vmatrix} l_{1,1}\lambda_1^{k_1-k} & l_{1,2}\lambda_2^{k_1-k} & \cdots & l_{1,l}\lambda_l^{k_1-k} \\ l_{2,1}\lambda_1^{k_2-k} & l_{2,2}\lambda_2^{k_2-k} & \cdots & l_{2,l}\lambda_l^{k_2-k} \\ \vdots & \vdots & \ddots & \vdots \\ l_{l,1}\lambda_1^{k_l-k} & l_{l,2}\lambda_2^{k_l-k} & \cdots & l_{l,l}\lambda_l^{k_l-k} \end{vmatrix}.$$

Then D is asymptotic to $\prod_{j=1}^{l} \lambda_j^{k_j-k}$, i.e.

$$\lim_{i_1,i_2,\ldots,i_l \to +\infty} \frac{D}{\prod_{j=1}^{l} \lambda_j^{k_j-k}} = 1. \tag{31}$$

Proof of Lemma 3: The determinant D has $l!$ terms, which have the form $sgn(\sigma) \prod_{j=1}^{l} l_{j,a_j} \lambda_{a_j}^{k_j-k}$. $\sigma = (a_1, a_2, \ldots, a_l)$ is a permutation of the set $\{1, 2, \ldots, l\}$ and $sgn(\sigma) = \pm 1$ is the sign of permutation. Rewrite (31) as

$$\frac{D}{\prod_{j=1}^{l} \lambda_j^{k_j-k}} = \sum_{\sigma} sgn(\sigma) \frac{\prod_{j=1}^{l} l_{j,a_j} \lambda_{a_j}^{k_j-k}}{\prod_{j=1}^{l} \lambda_j^{k_j-k}} = \sum_{\sigma} sgn(\sigma) \prod_{j=1}^{l} l_{j,a_j} \frac{\left(\prod_{j=1}^{l} \lambda_{a_j}\right)^{-i_1} \cdots \left(\prod_{j=l}^{l} \lambda_{a_j}\right)^{-i_l}}{\left(\prod_{j=1}^{l} \lambda_j\right)^{-i_1} \cdots \left(\prod_{j=l}^{l} \lambda_j\right)^{-i_l}}$$

$$= \sum_{\sigma} sgn(\sigma) \prod_{j=1}^{l} l_{j,a_j} \prod_{b=1}^{l} \left(\frac{\prod_{j=b}^{l} \lambda_{a_j}}{\prod_{j=b}^{l} \lambda_j} \right)^{-i_j}.$$

Now we can just analyze each term of the summation. Since $|\lambda_1| \geq \cdots \geq |\lambda_l|$, $|\prod_{j=b}^{l} \lambda_{a_j}| \geq |\prod_{j=b}^{l} \lambda_j|$. First consider that there exist some js such that $|\lambda_{a_j}| \neq |\lambda_j|$ and define j^* to be the largest, which means $|\lambda_{a_{j^*}}| \neq |\lambda_{j^*}|$ and $|\lambda_{a_j}| = |\lambda_j|$ for all j greater than j^*. Since $|\lambda_{j^*}|$ is the smallest among $|\lambda_1|, \ldots, |\lambda_j|$, we know that $|\lambda_{a_{j^*}}| > |\lambda_{j^*}|$. Thus,

$$\left| \frac{\prod_{j=j^*}^{l} \lambda_{a_j}}{\prod_{j=j^*}^{l} \lambda_j} \right| > 1,$$

$$\lim_{i_1, i_2, \ldots, i_l \to \infty} \left| \prod_{j=1}^{l} l_{j,a_j} \prod_{b=1}^{l} \left(\frac{\prod_{j=b}^{l} \lambda_{a_j}}{\prod_{j=b}^{l} \lambda_j} \right)^{-i_j} \right| \leq |\prod_{j=1}^{l} l_{a,a_j}| \lim_{i_{j^*} \to \infty} \left| \frac{\prod_{j=j^*}^{l} \lambda_{a_j}}{\prod_{j=j^*}^{l} \lambda_j} \right|^{-i_{j^*}} = 0.$$

Then consider that if for all j, $|\lambda_{a_j}| = |\lambda_j|$, but $(a_1, \ldots, a_l) \neq (1, 2, \ldots, l)$. Thus, there exists j^* such that $a_{j^*} \neq j^*$. Hence $l_{j^*, a_{j^*}} = 0$. Therefore, these terms are always 0.

The only term left is

$$sgn(\sigma) \prod_{j=1}^{l} l_{j,j} \prod_{b=1}^{l} \left(\frac{\prod_{j=b}^{l} \lambda_j}{\prod_{j=b}^{l} \lambda_j} \right)^{-i_j} = 1.$$

Thus, we can conclude that

$$\lim_{i_1, i_2, \ldots, i_l \to \infty} \frac{D}{\prod_{j=1}^{l} \lambda_j^{k_j - k}} = 1.$$

■

Because the system is non-degenerate, by Lemma 2, we know that there exist $L_1, L_2, \cdots, L_l$, such that $L_i C = [l_{i,1}, \ldots, l_{i,l}]$ is a row vector, $l_{i,i} = 1$ and $l_{i,a} = 0$ if $i \neq a$ and $|\lambda_i| = |\lambda_a|$.

Define matrices

$$U \triangleq \begin{bmatrix} l_{1,1}\lambda_1^{-i_1} & l_{1,2}\lambda_2^{-i_1} & \cdots & l_{1,l}\lambda_l^{-i_1} \\ l_{2,1}\lambda_1^{-i_2} & l_{2,2}\lambda_2^{-i_2} & \cdots & l_{2,l}\lambda_l^{-i_2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{l,1}\lambda_1^{-i_l} & l_{l,2}\lambda_2^{-i_l} & \cdots & l_{l,l}\lambda_l^{-i_l} \end{bmatrix}, O \triangleq U^{-1}. \tag{32}$$

Define $\alpha_5 = \max(\lambda_{max}(L_1^H L_1), \ldots, \lambda_{max}(L_l^H L_l))$. Thus, $L_i^H L_i \leq \alpha_5 I_m$, and

$$\begin{aligned} \sum_{j=1}^{l} A^{-i_j H} C^H C A^{-i_j} &\geq \sum_{j=1}^{l} \frac{1}{\alpha_5} A^{-i_j H} C^H L_j^H L_j C A^{-i_j} \\ &= \frac{1}{\alpha_5} \begin{bmatrix} A^{-i_1 H} C^H L_1^H & \cdots & A^{-i_l H} C^H L_1^H \end{bmatrix} \begin{bmatrix} L_1 C A^{-i_1} \\ \vdots \\ L_l C A^{-i_l} \end{bmatrix} = \frac{1}{\alpha_5} U^H U, \end{aligned} \tag{33}$$

and

$$\left(\sum_{j=1}^{l} A^{(k_j-k)H}C^HCA^{k_j-k}\right)^{-1} \leq \alpha_5\left(U^HU\right)^{-1} = \alpha_5 OO^H \leq \alpha_5 trace(OO^H)I_l$$
$$= \alpha_5\sum_{a,b} O_{a,b}(O^H)_{b,a}I_l = \alpha_5\sum_{a,b} O_{a,b}\times conj(O_{a,b})I_l = \alpha_5\sum_{a,b}|O_{a,b}|^2 I_l, \tag{34}$$

where $conj()$ means complex conjugation.

Now by Lemma 3, we can compute the cofactor matrix of U and hence $O = U^{-1}$. Define the minor $M_{a,b}$ of U as the $(l-1)\times(l-1)$ matrix that results from deleting row a and column b. Thus

$$O_{a,b} = \frac{(-1)^{a+b}\det(M_{b,a})}{\det(U)}. \tag{35}$$

By Lemma 3, we know that

$$\lim_{i_1,i_2,\ldots,i_l\to\infty}\frac{\det(U)}{\prod_{j=1}^{l}\lambda_j^{k_j-k}} = 1.$$

Since $M_{a,b}$ has the same structure as U, it is easy to show that

$$\det(M_{a,b}) \leq \rho_{a,b}\prod_{j=2}^{l}|\lambda_j^{k_{j-1}-k}|,$$

where $\rho_{a,b}$ is a constant. Thus,

$$\limsup_{i_1,\ldots,i_l\to\infty}\frac{\left(\sum_{j=1}^{l}A^{k_j-kH}C^HCA^{k_j-k}\right)^{-1}}{\prod_{j=1}^{l}|\lambda_j|^{2i_j}} \leq \limsup_{i_1,\ldots,i_l\to\infty}\frac{\alpha_5\sum_{a,b}|O_{a,b}|^2}{\prod_{j=1}^{l}|\lambda_j|^{2i_j}}I_l$$
$$= \limsup_{i_1,\ldots,i_l\to\infty}\alpha_5\left(\sum_{a,b}\left|\frac{\det(M_{a,b})}{\det(U)}\right|^2 / \prod_{j=1}^{l}|\lambda_j|^{2i_j}\right)I_l \tag{36}$$
$$\leq \alpha_5\left(\sum_{a,b}\rho_{a,b}^2\left|\frac{\prod_{j=2}^{l}|\lambda_j^{k_{j-1}-k}|}{\prod_{j=1}^{l}\lambda_j^{k_j-k}}\right|^2 / \prod_{j=1}^{l}|\lambda_j|^{2i_l}\right)I_l = \alpha_5\sum_{i,j}\rho_{i,j}^2 I_l.$$

Hence, there exists $\alpha_6 > 0$ such that

$$\left(\sum_{j=1}^{l}A^{k_j-kH}C^HCA^{k_j-k}\right)^{-1} \leq \alpha_6\prod_{j=1}^{l}|\lambda_j|^{2i_j}I_l,$$

Combining with Lemma 1 we can finish the proof.