arXiv:0910.4686v2 [math.PR] 3 Jun 2010

Moderate Deviations of the Random Riccati Equation

Soummya Kar and José M. F. Moura*

Abstract

We characterize the invariant filtering measures resulting from Kalman filtering with intermittent observations ([1]), where the observation arrival is modeled as a Bernoulli process. In [2], it was shown that there exists a $\overline{\gamma}^{\text{sb}} > 0$ such that for every observation packet arrival probability $\overline{\gamma}$, $\overline{\gamma} > \overline{\gamma}^{\text{sb}} > 0$, the sequence of random conditional error covariance matrices converges in distribution to a unique invariant distribution $\mu^{\overline{\gamma}}$ (independent of the filter initialization.) In this paper, we prove that, for controllable and observable systems, $\overline{\gamma}^{\text{sb}} = 0$ and that, as $\overline{\gamma} \uparrow 1$, the family $\{\mu^{\overline{\gamma}}\}_{\overline{\gamma}>0}$ of invariant distributions satisfies a moderate deviations principle (MDP) with a good rate function I. The rate function I is explicitly identified. In particular, our results show: (1) as $\overline{\gamma} \uparrow 1$, the family $\{\mu^{\overline{\gamma}}\}$ converges weakly (in distribution) to the Dirac measure δ_{P^*}, where P^* is the fixed point of the discrete time Riccati operator; (2) the probability of a rare event (an event bounded away from P^*) under $\mu^{\overline{\gamma}}$ decays to zero as a power law of $(1-\overline{\gamma})$ as $\overline{\gamma} \uparrow 1$. The best exponent of such a power law decay is explicitly obtained by solving a deterministic variational problem involving the MDP rate function I. These results offer a complete characterization of the family of invariant distributions $\{\mu^{\overline{\gamma}}\}_{\overline{\gamma}>0}$.We provide computationally efficient methods for solving the variational problems in question, leading to efficient estimates of probabilities under the invariant measures. The analytical techniques developed in this paper are fairly general and applicable to the analysis of a broader class of iterated function systems. Several intermediate results obtained in the process are of independent interest.

1. Introduction

A. Background and Motivation

Kalman filtering with non-classical information pattern has received significant attention in the control and signal processing literature. There has been renewed interest, motivated by increasing real-time networked systems applications. Such networks operate under constrained resources with lack of supervised control centers leading to inherent sources of randomness in the information pattern. For reliable system operation, it is of interest to understand the asymptotic properties of such systems like stability and ergodicity. In [2], we studied this problem in the context of Kalman filtering with intermittent observations ([1].) The results in [2] establish an interesting dichotomy for the filtering error process and show, in particular, that stochastic boundedness of the sequence of conditional error covariance matrices (generated by the discrete time random Riccati equation (RRE)) is necessary and sufficient for its ergodicity. In other words, we showed the existence of a critical probability, $\overline{\gamma}^{\text{sb}}$, such that, if the observation packet arrival probability $\overline{\gamma}$ is greater than $\overline{\gamma}^{\text{sb}}$, the sequence of random error covariance matrices converges weakly (in distribution) to a unique invariant distribution $\mu^{\overline{\gamma}}$. We note here that stochastic boundedness is a much weaker condition than moment stability, and, as shown in [2], convergence to a unique invariant distribution is possible under a packet arrival probability for which moment stability does not hold. In this context, we further note, that our work ([2]) provides a sample-path analysis of the RRE, in contrast to moment stability analysis, as is done conventionally in the literature (see, for example, [3], [4], [5], [6], [7], [8], [9], [10], [11] and also [2] for a detailed review of the literature.)

The authors are with the Dept. Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA (e-mail: soummyak@andrew.cmu.edu, moura@ece.cmu.edu, ph: (412)268-6341, fax: (412)268-3890.)

Work supported by NSF under grants # ECS-0225449 and # CNS-0428404.

To summarize, the results in [2] showed the existence and uniqueness of an attracting invariant measure $\mu^{\overline{\gamma}}$ for the RRE, for every $\overline{\gamma} > \overline{\gamma}^{\text{sb}}$, to which the conditional error covariance matrices converge weakly when operated at packet arrival probability $\overline{\gamma}$. In this paper, we prove that, for observable and controllable systems, $\overline{\gamma}^{\text{sb}} = 0$, and hence $\mu^{\overline{\gamma}}$ exists and is unique for every $\overline{\gamma} > 0$ for such systems[1]. The main goal of this paper is to undertake the highly nontrivial problem of characterizing the resulting invariant measures $\mu^{\overline{\gamma}}$. In the non-classical information case, characterization of the steady-state error covariance distribution $\mu^{\overline{\gamma}}$ is as important as characterizing the deterministic fixed point of the Riccati equation in the classical case, as derived by Kalman [12] for discrete time and by Kalman and Bucy [13] for continuous time.

We detail the key contributions of this paper. We show the following for observable and controllable systems.

1. **Stochastic boundedness: $\overline{\gamma}^{\text{sb}} = 0$.** We prove the result stated in Theorem 9 in [2], by proving that, for controllable and observable systems, $\overline{\gamma}^{\text{sb}} = 0$, and so, for any non-zero observation arrival probability $\overline{\gamma}$, the conditional error covariance process is ergodic, with the unique attracting measure $\mu^{\overline{\gamma}}$.
2. **Moderate deviation principle (MDP).** We show that the family of invariant distributions $\left\{\mu^{\overline{\gamma}}\right\}_{\overline{\gamma}>0}$ satisfies a moderate deviations principle (MDP) with good rate function I as $\overline{\gamma} \uparrow 1$. An immediate consequence (which is rather intuitive but not natural) is that, as $\overline{\gamma} \uparrow 1$, the family of invariant measures $\left\{\mu^{\overline{\gamma}}\right\}$ converges weakly to the Dirac mass δ_{P^*}, where P^* is the unique fixed point of the deterministic Riccati equation.
3. **Probability of rare events.** The MDP implies that the probabilities of 'rare events' (events bounded away from P^*) decay to zero as $\overline{\gamma} \uparrow 1$. A natural question of practical and theoretical interest is the rate at which the probability of such a rare event goes to zero. We show that the probability of such rare events decays as a power law of $(1-\overline{\gamma})$ as $\overline{\gamma} \uparrow 1$.
4. **Best decay: Variational problem.** The best exponent for the power law decay of the probability of rare events depends on the rare event of interest; it can be explicitly characterized in terms of the rate function $I(\cdot)$. Formally, we have the following MDP asymptotics:

$$\mu^{\overline{\gamma}}(\Gamma) \sim (1-\overline{\gamma})^{\inf_{x\in\Gamma} I(x)} \tag{1}$$

(this notation is made precise in the paper.) Thus, the exact decay asymptotics of a rare event is obtained by solving a variational problem involving the rate function I. Since, the above MDP asymptotics holds for every Borel set Γ, our result characterizes completely the family of invariant measures $\left\{\mu^{\overline{\gamma}}\right\}$.
5. **Estimating the probability of rare events.** The estimation of probabilities of rare events reduces to solving deterministic variational problems; this not only characterizes the decay rate of rare events but also gives insight into how such events occur. In Section 8, we show several techniques that can be employed to solve these variational problems efficiently. We emphasize that our analysis of reducing the problem of estimating probabilities of interest to solving variational problems efficiently is much more definitive and relevant than numerically estimating the invariant distributions. A naive numerical approach of simulating the distributions $\left\{\mu^{\overline{\gamma}}\right\}$ as $\overline{\gamma} \uparrow 1$ becomes meaningless as the rare events of interest become increasingly difficult to observe as $\overline{\gamma} \uparrow 1$ (see also Section 9.) One may take recourse to sophisticated simulation techniques like importance sampling (see, for example, [14]), but such approaches require characterization of the distributions in question, which is addressed in this paper.

More broadly, the techniques developed in this paper are fairly general and go beyond the setting of Kalman

[1]The fact, that $\overline{\gamma}^{\text{sb}} = 0$, was proved for systems with invertible observation matrices in [2]. The proof for general observable systems is provided in Appendix B of the present paper.

filtering with intermittent observations. There is a key difference in the MDP arguments used here and conventional methods for analyzing the moderate (or large) deviations for stationary measures of Markov processes, where it is generally assumed that the underlying Markov process is positive recurrent and moderate deviations of stationary measures then follow from that of finite dimensional distributions (see, for example, [15],[16].) However, the Markov processes governing the RRE are not, in general, positive recurrent, as is the case for a large class of iterated function systems ([17].) Our analysis proceeds by studying the probability measures induced on the space of random function compositions (strings) and developing its topological properties, as detailed in the paper. Several intermediate results obtained in the process are of independent interest and follow under more general assumptions. Our tools are applicable to the analysis of more complex networked control systems (see, for example, [18]) and hybrid or switched systems.

We summarize the organization of the paper. Subsection 1-B presents notation and preliminaries on moderate deviations. Subsection 2-A sets up the problem and prior work is briefly reviewed in Subsection 2-B. Several key approximation results are presented in Section 3, whereas the main results of this paper are stated and discussed in Section 4. MDP for finite dimensional distributions of the RRE sequence is analyzed in Section 5, whereas Section 6 systematically carries out the steps required to obtain the main results on MDP for stationary distributions, which is completed in Section 7. In Section 8, we present efficient ways to solve the variational problems involving the rate function. Numerical studies justifying the theoretical results for a scalar system arepresented in Section 9. Section 10 concludes the paper.

B. Notation and Preliminaries

Denote by: $\mathbb{R}$, the reals; $\mathbb{R}^M$, the M-dimensional Euclidean space; $\mathbb{T}$, the integers; $\mathbb{T}_+$, the non-negative integers; $\mathbb{N}$, the natural numbers; and $\mathcal{X}$, a generic space. For a subset $B \subset \mathcal{X}$, $\mathbb{I}_B : \mathcal{X} \longmapsto \{0,1\}$ is the indicator function, which is 1 when the argument is in B and zero otherwise; and $\mathrm{id}_{\mathcal{X}}$ is the identity function on $\mathcal{X}$. A metric space $\mathcal{X}$ with metric $d_{\mathcal{X}}$ is denoted by the pair $(\mathcal{X}, d_{\mathcal{X}})$. The corresponding Borel algebra is denoted by $\mathcal{B}(\mathcal{X})$. For $x \in \mathcal{X}$, the open ball of radius $\varepsilon > 0$ centered at x is denoted by $B_\varepsilon(x)$, i.e., $B_\varepsilon(x) = \{y \in \mathcal{X} \mid d_{\mathcal{X}}(y,x) < \varepsilon\}$. The closure of $B_\varepsilon(x)$ is the closed ball of radius $\varepsilon > 0$ centered at x and is denoted by $\overline{B}_\varepsilon(x)$. For any set $\Gamma \subset \mathcal{X}$, the open ε-neighborhood of Γ is given by

$$\Gamma_\varepsilon = \left\{ y \in \mathcal{X} \mid \inf_{x \in \Gamma} d_{\mathcal{X}}(y,x) < \varepsilon \right\} \tag{2}$$

It can be shown that Γ_ε is an open set. Similarly, the closed ε-neighborhood of Γ is given by

$$\overline{\Gamma_\varepsilon} = \left\{ y \in \mathcal{X} \mid \inf_{x \in \Gamma} d_{\mathcal{X}}(y,x) \leq \varepsilon \right\} \tag{3}$$

which is a closed set. For a set $\Gamma \subset \mathcal{X}$, we denote by Γ° and $\overline{\Gamma}$ its interior and closure respectively.

The Banach space of symmetric matrices

Let $\mathbb{S}^N$ denote the separable Banach space of symmetric $N \times N$ matrices, equipped with the induced 2-norm. The subset $\mathbb{S}^N_+$ of positive semidefinite matrices is a closed, convex, solid, normal, minihedral cone in $\mathbb{S}^N$, with non-empty interior $\mathbb{S}^N_{++}$, the set of positive definite matrices. The cone $\mathbb{S}^N_+$ induces a partial order in $\mathbb{S}^N$. For $X, Y \in \mathbb{S}^N$, we write $X \preceq Y$ ($Y \succeq X$) to denote $Y - X \in \mathbb{S}^N_+$; $X \prec Y$ to denote $X \preceq Y$ and $X \neq Y$; $X \ll Y$ ($Y \gg X$) to denote $Y - X \in \mathbb{S}^N_{++}$.

Limit Notations Let $h : \mathbb{R} \longmapsto \mathbb{R}$ be a measurable function.

The notation $\lim_{z \to x} f(z) = y$ implies that for every sequence $\{z_n\}_{n \in \mathbb{N}}$ in $\mathbb{R}$ with $\lim_{n \to \infty} |z_n - x| = 0$, we have $\lim_{n \to \infty} |f(z_n) - y| = 0$. The notation $\lim_{z \uparrow x} f(z) = y$ implies that for every sequence $\{z_n\}_{n \in \mathbb{N}}$ in $\mathbb{R}$ with $z_n < x$

and $\lim_{n\to\infty}|z_n - x| = 0$, we have $\lim_{n\to\infty}|f(z_n) - y| = 0$. The notations $\downarrow$ and $\uparrow$ have similar implications when working with limit inferiors and superiors.

Probability measures on metric spaces: Let: $(\mathcal{X}, d_{\mathcal{X}})$ a complete separable metric space $\mathcal{X}$ with metric $d_{\mathcal{X}}$; $\mathbb{B}(\mathcal{X})$ its Borel algebra; $B(\mathcal{X})$ the Banach space of real-valued bounded functions on $\mathcal{X}$, equipped with the sup-norm, i.e., $f \in B(\mathcal{X})$, $\|f\| = \sup_{x\in\mathcal{X}}|f(x)|$; and $C_b(\mathcal{X})$ the subspace of $B(\mathcal{X})$ of continuous functions. Let $\mathcal{P}(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. For $\mu \in \mathcal{M}(\mathcal{X})$, we define the support of μ, supp(μ), by

$$\text{supp}(\mu) = \{x \in \mathcal{X} \mid \mu(B_\varepsilon(x)) > 0, \ \forall \varepsilon > 0\} \tag{4}$$

It follows that supp(μ) is a closed set. The sequence $\{\mu_t\}_{t\in\mathbb{T}_+}$ in $\mathcal{P}(\mathcal{X})$ converges weakly to $\mu \in \mathcal{P}(\mathcal{X})$ if

$$\lim_{t\to\infty} < f, \mu_t > = < f, \mu >, \quad \forall\ f \in C_b(\mathcal{X}) \tag{5}$$

Weak convergence is denoted by $\mu_t \Longrightarrow \mu$ and is also referred to as convergence in distribution. The weak topology on $\mathcal{P}(\mathcal{X})$ generated by weak convergence can be metrized. In particular, e.g., [19], one has the Prohorov metric d_p on $\mathcal{P}(\mathcal{X})$, such that the metric space $(\mathcal{P}(\mathcal{X}), d_p)$ is complete, separable, and a sequence $\{\mu_t\}_{t\in\mathbb{T}_+}$ in $\mathcal{P}(\mathcal{X})$ converges weakly to μ in $\mathcal{P}(\mathcal{X})$ *iff* $\lim_{t\to\infty} d_p(\mu_t, \mu) = 0$. The distance between two probability measures μ_1, μ_2 in $\mathcal{P}(\mathcal{X})$ is computed as:

$$d_P(\mu_1, \mu_2) = \inf\{\varepsilon > 0 \mid \mu_1(\mathcal{F}) \le \mu_2(\mathcal{F}_\varepsilon) + \varepsilon, \quad \forall \text{ closed set } \mathcal{F}\} \tag{6}$$

Moderate Deviations:

Definition 1.1 Let $\{\mu^{\overline{\gamma}}\}$ be a family of probability measures on the complete separable metric space $(\mathcal{X}, d_{\mathcal{X}})$ indexed by the real-valued parameter $\overline{\gamma}$ taking values in $(0, 1)$. Let $h : (0, 1) \longmapsto \mathbb{R}_+$ be a non-decreasing function on $(0, 1)$ with

$$\lim_{\overline{\gamma}\uparrow 1} h(\overline{\gamma}) = \infty \tag{7}$$

Let $I : \mathcal{X} \longmapsto \overline{\mathbb{R}}_+$ be an extended valued lower semicontinuous function. The family $\{\mu^{\overline{\gamma}}\}$ is said to satisfy a moderate deviations principle (MDP) with rate function $I(\cdot)$ at scale $h(\overline{\gamma})$ as $\overline{\gamma} \uparrow 1$ if the following holds:

$$\liminf_{\overline{\gamma}\uparrow 1} \frac{1}{h(\overline{\gamma})} \ln \mu^{\overline{\gamma}}(\mathcal{O}) \ge -\inf_{x\in\mathcal{O}} I(x), \quad \text{for every open set } \mathcal{O} \in \mathcal{X} \tag{8}$$

$$\limsup_{\overline{\gamma}\uparrow 1} \frac{1}{h(\overline{\gamma})} \ln \mu^{\overline{\gamma}}(\mathcal{F}) \le -\inf_{x\in\mathcal{F}} I(x), \quad \text{for every closed set } \mathcal{F} \in \mathcal{X} \tag{9}$$

The function $I(\cdot)$ is called the MDP rate function. The lower semicontinuity implies that the level sets of $I(\cdot)$, i.e., sets of the form $\{x \in \mathcal{X} \mid I(x) \le \alpha\}$ for every $\alpha \in \mathbb{R}_+$, are closed. If in addition, the levels sets are compact (for every α), $I(\cdot)$ is said to be a good rate function, and the corresponding family $\{\mu^{\overline{\gamma}}\}$ is said to satisfy an MDP with good rate function $I(\cdot)$.

It can be shown that the MDP, as stated in (8)-(9), is equivalent to the following:

$$-\inf_{x\in\Gamma^\circ} I(x) \le \liminf_{\overline{\gamma}\uparrow 1} \frac{1}{h(\overline{\gamma})} \ln \mu^{\overline{\gamma}}(\Gamma) \le \limsup_{\overline{\gamma}\uparrow 1} \frac{1}{h(\overline{\gamma})} \ln \mu^{\overline{\gamma}}(\Gamma) \le -\inf_{x\in\overline{\Gamma}} I(x) \tag{10}$$

for every measurable set Γ. In other words, (10) holds *iff* (8)-(9) hold.

The above formulation of MDP is similar in spirit to the theory of large deviations principle (LDP). In fact, in the above definition, if the scale function $h(\cdot)$ is a polynomial in $\overline{\gamma}$, the family $\{\mu^{\overline{\gamma}}\}$ is said to satisfy an LDP (see,

for example, [15], [20].) This conceptual similarity is manifested in some of the proof techniques developed in the paper having parallels with their counterparts in the theory of LDP.

Before interpreting the consequences of an MDP as defined above, we consider the notion of a *rare event*, which is the central motivation to all MDP (and LDP):

Definition 1.2 (Rare Event) : A set $\Gamma \subset \mathcal{B}(\mathcal{X})$ is called a rare event with respect to the family $\{\mu^{\overline{\gamma}}\}$ of probability measures, if $\lim_{\overline{\gamma}\uparrow 1} \mu^{\overline{\gamma}}(\Gamma) = 0$. In other words, the event Γ becomes increasingly difficult to observe (i.e., it becomes rare) as $\overline{\gamma} \uparrow 1$.

Once a rare event Γ is identified, the next natural question is the rate at which its probability goes to zero under $\mu^{\overline{\gamma}}$ as $\overline{\gamma} \uparrow 1$. This is answered by an MDP, which also gives a complete characterization of the family as $\overline{\gamma} \uparrow 1$. Indeed, from Def. 1.1 it is not hard to see that, if the family $\{\mu^{\overline{\gamma}}\}$ satisfies an MDP, we have for every measurable set $\Gamma \in \mathcal{X}$:

$$c_1(\overline{\gamma}) e^{-h(\overline{\gamma}) \inf_{x\in\Gamma^\circ} I(x)} \leq \mu^{\overline{\gamma}}(\Gamma) \leq c_2(\overline{\gamma}) e^{-h(\overline{\gamma}) \inf_{x\in\overline{\Gamma}} I(x)} \tag{11}$$

where $c_1, c_2 : (0,1) \longmapsto \mathbb{R}_+$ are functions, such that, $\lim_{\overline{\gamma}\uparrow 1} c_i(\overline{\gamma}) = 1, \quad i = 1, 2$. For brevity, we subsequently use the notation

$$\mu^{\overline{\gamma}}(\Gamma) \sim e^{-h(\overline{\gamma}) \inf_{x\in\Gamma} I(x)} \tag{12}$$

as a short form of (11).

Now assume Γ is a rare event and, to avoid unnecessary technicalities, also assume that Γ is an I-continuity set, i.e., $\inf_{x\in\Gamma^\circ} = \inf_{x\in\overline{\Gamma}} = \inf_{x\in\Gamma}$. Then, by (11) we must have $\inf_{x\in\Gamma} > 0$ (so that the probabilities decay to zero.) Thus, (11) implies that the probability of the rare event Γ decays exponentially at a scale $h(\overline{\gamma})$ to zero, $\inf_{x\in\Gamma} I(x)$ being best exponent (or rate) of decay. Such a characterization of the best decay rate of a rare event is extremely important in system analysis and, as will be seen in the paper, offers considerable insight into system design apart from providing a complete characterization of the measures $\mu^{\overline{\gamma}}$.

2. Problem Formulation

We split the present section into two subsections, Subsection 2-A briefly summarizing the model of Kalman filtering with intermittent observations, while in Subsection 2-B we review some results from [2] on weak convergence of the random error covariance matrices resulting from the above filtering model.

A. Setup

We start by reviewing the model of Kalman filtering with intermittent observations in [1]. Let

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_t \tag{13}$$

$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t \tag{14}$$

Here $\mathbf{x}_t \in \mathbb{R}^N$ is the signal (state) vector, $\mathbf{y}_t \in \mathbb{R}^M$ is the observation vector, $\mathbf{w}_t \in \mathbb{R}^N$ and $\mathbf{v}_t \in \mathbb{R}^M$ are Gaussian random vectors with zero mean and covariance matrices Q and R, respectively. The sequences $\{\mathbf{w}_t\}_{t\in\mathbb{T}_+}$ and $\{\mathbf{v}_t\}_{t\in\mathbb{T}_+}$ are uncorrelated and mutually independent. Also, assume that the initial state $\mathbf{x}_0$ is a zero-mean Gaussian vector with covariance P_0. Unless otherwise stated, we use the following standing assumption throughout the paper:

Assumption (E): The pair (C, A) is observable and Q, R are positive definite. The assumption $Q \gg 0$ implies the controllability of the pair $(A, Q^{1/2})$. The main results of the paper require all these assumptions, but several

intermediate results of independent interest hold under less stringent assumptions. In that case, they are noted explicitly.

The m.m.s.e. predictor $\widehat{\mathbf{x}}_{t|t-1}$ of the signal vector $\mathbf{x}_t$ given the observations $\{\mathbf{y}_s\}_{0\leq s<t}$ is the conditional mean. It is recursively implemented by the Kalman filter. The sequence of conditional prediction error covariances, $\{P_t\}_{t\in\mathbb{T}_+}$, is then given by

$$P_t = \mathbb{E}\left[\left(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t-1}\right)\left(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t-1}\right)^T \mid \{\mathbf{y}(s)\}_{0\leq s<t}\right] \tag{15}$$

$$P_{t+1} = AP_tA^T+Q-AP_tC^T\left(CP_tC^T+R\right)^{-1}CP_tA^T \tag{16}$$

Under the hypothesis of controllability of the pair $(A,Q^{1/2})$ and observability of the pair (C,A), the deterministic sequence $\{P_t\}_{t\in\mathbb{T}_+}$ converges to a unique value P^* (which is a fixed point of the algebraic Riccati equation (16)) from any initial condition P_0. [12].

This corresponds to the classical perfect observation scenario, where the estimator has complete knowledge of the observation packet $\mathbf{y}_t$ at every time t. With intermittent observations, the observation packets are dropped randomly (across the communication channel to the estimator), and the estimator receives observations at random times. We study the intermittent observation model considered in [1], where the channel randomness is modeled by a sequence $\{\gamma_t\}_{t\in\mathbb{T}_+}$ of i.i.d. Bernoulli random variables with mean $\overline{\gamma}$ (note, $\overline{\gamma}$ then denotes the arrival probability.) Here, $\gamma_t=1$ corresponds to the arrival of the observation packet $\mathbf{y}_t$ at time t to the estimator, whereas a packet dropout corresponds to $\gamma_t=0$. Denote by $\widetilde{\mathbf{y}}_t$ the pair $\widetilde{\mathbf{y}}_t=\left(\mathbf{y}_t\mathbb{I}_{(\gamma_t=1)},\gamma_t\right)$. Under the TCP packet acknowledgement protocol in [1] (the estimator knows at each time whether the observation packet arrived or not), the m.m.s.e. predictor of the signal is given by:

$$\widehat{\mathbf{x}}_{t|t-1}=\mathbb{E}\left[\mathbf{x}_t\,\middle|\,\{\widetilde{\mathbf{y}}_s\}_{0\leq s<t}\right] \tag{17}$$

A modified form of the Kalman filter giving a recursive implementation of the estimator in (17) is in [1]. The sequence of conditional prediction error covariance matrices, $\{P_t\}_{t\in\mathbb{T}_+}$, is updated according to the following random Riccati equation (RRE):

$$P_t = \mathbb{E}\left[\left(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t-1}\right)\left(\mathbf{x}_t-\widehat{\mathbf{x}}_{t|t-1}\right)^T \mid \{\widetilde{\mathbf{y}}(s)\}_{0\leq s<t}\right] \tag{18}$$

$$P_{t+1} = AP_tA^T+Q-\gamma_tAP_tC^T\left(CP_tC^T+R\right)^{-1}CP_tA^T \tag{19}$$

A convenient representation of P_t is obtained by defining the functions $f_0,f_1:\mathbb{S}_+^N\longmapsto\mathbb{S}_+^N$ as[2]:

$$f_0(X)=AXA^T+Q,\quad \forall X\in\mathbb{S}_+^N \tag{20}$$

$$f_1(X)=AXA^T+Q-\gamma_tAXC^T\left(CXC^T+R\right)^{-1}CXA^T,\quad \forall X\in\mathbb{S}_+^N \tag{21}$$

We then have for all $t\geq 1$

$$P_t=f_{\gamma_{t-1}}\circ f_{\gamma_{t-2}}\circ\cdots\circ f_{\gamma_0}(P_0) \tag{22}$$

Unlike the classical case, the sequence $\{P_t\}_{t\in\mathbb{T}_+}$ is now random (because of its dependence on the random sequence $\{\gamma_t\}_{t\in\mathbb{T}_+}$.) Thus, for each t, P_t is a random element of S_+^N, and we denote by $\mu_t^{\overline{\gamma},P_0}$ its distribution (the measure it induces on S_+^N.) The superscripts $\overline{\gamma}$, P_0 emphasize the dependence of $\mu_t^{\overline{\gamma},P_0}$ on the packet arrival probability and the initial condition. We often use the notations $\mathbb{P}^{\overline{\gamma},P_0},\mathbb{E}^{\overline{\gamma},P_0}$ to denote probability and expectation operators

[2] f_0 corresponds to the Lyapunov operator, whereas f_1 is the Riccati operator.

respectively, when the system is operated with observation arrival probability $\overline{\gamma}$ and initial covariance P_0.

B. Prior work

The following extends Theorems 9, 10 in [2] on the weak convergence of the RRE sequence.

Theorem 2.1 Let assumption **(E.1)** hold. Then,

- For each $\overline{\gamma} > 0$, there exists a unique invariant distribution $\mu^{\overline{\gamma}}$ s.t. the sequence $\{P_t\}_{t\in\mathbb{T}_+}$ (or sequence $\left\{\mu_t^{\overline{\gamma},P_0}\right\}_{t\in\mathbb{T}_+}$ of measures) converges weakly to $\mu^{\overline{\gamma}}$ from any initial condition $P_0 \in \mathbb{S}_+^N$.

- Define the set $\mathcal{S} \subset \mathbb{S}_+^N$ by

$$\mathcal{S} = \{f_{i_1} \circ f_{i_2} \circ \cdots \circ f_{i_s}(P^*) \mid i_r \in \{0,1\},\ 1 \le r \le s,\ s \in \mathbb{T}_+\} \tag{23}$$

  Then[3], if $0 < \overline{\gamma} < 1$,

$$\text{supp}\left(\mu^{\overline{\gamma}}\right) = \text{cl}(\mathcal{S}) \tag{24}$$

  where $\text{cl}(\mathcal{S})$ denotes the topological closure of $\mathcal{S}$ in $\mathbb{S}_+^N$ and supp denotes the support of a probability measure ([21]). In particular, we have

$$\mu^{\overline{\gamma}}\left(\left\{Y \in \mathbb{S}_+^N \mid Y \succeq P^*\right\}\right) = 1 \tag{25}$$

The proof is in Appendix B. For a detailed discussion of the above results, the reader is referred to [2].

3. Some key approximation results

In this section we present some results on random compositions of Lyapunov and Riccati operators leading to the RRE sequence (Subsection 3-A.) In Subsection 3-B, we present some key approximation results that are of independent interest and establish several useful properties of the RRE and the classical Riccati operator.

A. Preliminary Results

The RRE sequence is an iterated function system (see, for example, [17]) comprising of random compositions of Lyapunov and Riccati operators. To understand the system, we study the behavior of such random function compositions, where not only the numerical value of the composition is important, but also the composition pattern is relevant. To formalize this study, we start with the following definitions:

Definition 3.1 (String) : Let $P_0 \in \mathbb{S}_+$. A string $\mathcal{R}$ with initial state P_0 and length $n \in \mathbb{T}_+$ is a $(n+1)$-tuple of the form:

$$\mathcal{R} = (f_{i_1}, f_{i_2}, \cdots f_{i_n}, P_0), \quad i_1, \cdots, i_n \in \{0,1\} \tag{26}$$

where f_0 and f_1 correspond to the Lyapunov and Riccati updates in eqns. (20,21). The length of a string $\mathcal{R}$ is denoted by $\text{len}(\mathcal{R})$. The set of all possible strings is denoted by $\overline{\mathcal{S}}$.

Remark 3.2 Note that a string $\mathcal{R}$ can be of length 0; then it is represented as a 1-tuple, consisting of only the initial condition. We introduce notation here. Let $t_1, t_2, \cdots, t_l$ be non-negative integers, such that, $\sum_{i=1}^{l} t_i = n$

[3]In the definition of $\mathcal{S}$ ((23)), s can take the value 0, implying $P^* \in \mathcal{S}$.

and $i_j^k \in \{0,1\}$ for $1 \leq j \leq t_k, \quad 1 \leq k \leq l$, such that, for all k, $i_j^k = i_1^k, \ 1 \leq j \leq t_k$. Let $\mathcal{R}$ be a string of length n of the form:

$$\mathcal{R} = \left(f_{i_1^1}, \cdots, f_{i_{t_1}^1}, \cdots, f_{i_1^2}, \cdots, f_{i_{t_2}^2}, \cdots, f_{i_1^l}, \cdots, f_{i_{t_l}^l}, P_0\right) \tag{27}$$

where the indices i_j^k satisfy the relations above. For brevity, we will write $\mathcal{R}$ as:

$$\mathcal{R} = \left(f_{i_1^1}^{t_1}, f_{i_1^2}^{t_2}, \cdots, f_{i_1^l}^{t_l}, P_0\right) \tag{28}$$

For example, the string $(f_0, f_1, f_1, f_1, f_0, f_0, P_0)$ is written concisely as $\left(f_0, f_1^3, f_0^2, P_0\right)$.

Definition 3.3 (Numerical Value of a String) : To every string $\mathcal{R}$ is associated its numerical value, denoted by $\mathcal{N}(\mathcal{R})$, which is the numerical evaluation of the function composition on the initial state P_0, i.e., for $\mathcal{R}$ of the form $\mathcal{R} = (f_{i_1}, f_{i_2}, \cdots f_{i_n}, P_0), \quad i_1, \cdots, i_n \in \{0,1\}$, we have[4]

$$\mathcal{N}(\mathcal{R}) = f_{i_1} \circ f_{i_2} \circ \cdots \circ f_{i_n}(P_0) \tag{29}$$

Thus, the numerical value can be viewed as a function $\mathcal{N}(\cdot)$ from the space $\overline{\mathcal{S}}$ of strings to $\mathbb{S}_+^N$. We abuse notation by denoting $\mathcal{N}(\overline{\mathcal{S}})$ to be the set of numerical values attainable, i.e.,

$$\mathcal{N}(\overline{\mathcal{S}}) = \left\{\mathcal{N}(\mathcal{R}) \mid \mathcal{R} \in \overline{\mathcal{S}}\right\} \tag{30}$$

Remark 3.4 Note the difference between a string and its numerical value. Two strings are equal *iff* they comprise of the same *order* of function compositions applied to the same initial state. In particular, two strings can be different, even if they evaluate to the same numerical value.

Definition 3.5 (Concatenated Strings) : Let $n \in \mathbb{N}$ and $t_1 \leq t_2 \leq \cdots \leq t_n \in \mathbb{T}_+$. Also, if $t_n \geq 1$, choose $i_1, i_2, \cdots, i_{t_n} \in \{0,1\}$. Then

$$\overline{\mathcal{R}} = \left(\left(f_{i_{t_1}}, f_{i_{t_1-1}} \cdots, f_{i_1}, P_0\right), \left(f_{i_{t_2}}, \cdots, f_{i_{t_1}}, \cdots, f_{i_1}, P_0\right), \cdots, \left(f_{i_{t_n}}, \cdots, f_{i_1}, P_0\right)\right) \tag{31}$$

is a concatenated string of block length n with initial state P_0[5].

We similarly define the numerical value of such a concatenated string $\overline{\mathcal{R}}$ by

$$\mathcal{N}(\overline{\mathcal{R}}) = \left(f_{i_{t_1}} \circ \cdots \circ f_{i_1}(P_0), f_{i_{t_2}} \circ \cdots \circ f_{i_{t_1}} \circ \cdots f_{i_1}(P_0), \cdots, f_{i_{t_n}} \circ \cdots \circ f_{i_1}(P_0)\right) \tag{32}$$

and note that $\mathcal{N}(\overline{\mathcal{R}}) \in \bigotimes_{i=1}^n \mathbb{S}_+^N$.

For fixed $P_0, n, t_1 \leq t_2 \leq \cdots \leq t_n$, the set of such concatenated strings is denoted by $\mathcal{S}_{t_1,\cdots,t_n}^{P_0}$. The corresponding set of numerical values is denoted by $\mathcal{N}(\mathcal{S}_{t_1,\cdots,t_n}^{P_0})$.

Finally, for $\mathbf{X} \in \bigotimes_{i=1}^n \mathbb{S}_+^N$, the set $\mathcal{S}_{t_1,\cdots,t_n}^{P_0}(\mathbf{X}) \subset \mathcal{S}_{t_1,\cdots,t_n}^{P_0}$ consists of all strings with numerical value $\mathbf{X}$, i.e.,

$$\mathcal{S}_{t_1,\cdots,t_n}^{P_0}(\mathbf{X}) = \left\{\overline{\mathcal{R}} \in \mathcal{S}_{t_1,\cdots,t_n}^{P_0} \mid \mathcal{N}\left(\overline{\mathcal{R}}\right) = \mathbf{X}\right\} \tag{33}$$

A rigorous algebra of such strings can be developed, which we undertake elsewhere. In the following, we present some important properties of strings to be used later (see Appendix A for a proof):

[4]For function compositions, we adopt a similar notation to that of strings, namely, for example, we denote the composition $f_0 \circ f_1 \circ f_1 \circ f_1 \circ f_0 \circ f_0(P_0)$ by $f_0 \circ f_1^3 \circ f_0^2(P_0)$.

[5]We again adopt the convention that the first block is simply P_0 if $t_1 = 0$.

Proposition 3.6 (i) For $s,t \in \mathbb{T}_+$ and $s \leq t$, we have, $\mathcal{N}\left(\mathcal{S}_s^{P^*}\right) \subset \mathcal{N}\left(\mathcal{S}_t^{P^*}\right)$. In particular, if for some $X \in \mathbb{S}_+^N$,$t_0 \in \mathbb{T}_+$ and $i_1, \cdots, i_{t_0} \in \{0,1\}$, the string $\mathcal{R} = \left(f_{i_1}, \cdots, f_{i_{t_0}}, P^*\right)$ belongs to $\mathcal{S}_{t_0}^{P^*}(X)$, we have

$$\left(f_{i_1}, \cdots, f_{i_{t_0}}, f_1^{t-t_0}, P^*\right) \in \mathcal{S}_t^{P^*}(X) \subset \mathcal{S}^{P^*}(X), \quad \forall t \geq t_0 \tag{34}$$

(ii) Fix $n \in \mathbb{N}$ and $t_1 < \cdots < t_n \in \mathbb{T}_+$. We then have, for $0 \leq \overline{\gamma} \leq 1$,

$$\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1}, \cdots, P_{t_n}) \in \mathcal{N}\left(\mathcal{S}_{t_1,\cdots,t_n}^{P_0}\right)\right) = 1 \tag{35}$$

(iii) Let $t \in \mathbb{T}_+$ and $\mathcal{R} \in \mathcal{S}_t^{P_0} = (f_{i_1}, \cdots, f_{i_t}, P_0)$ be a string. Then, there exists $\alpha_{P^0} \in \mathbb{R}_+$, depending on P^0, such that,

$$f_0^{\pi(\mathcal{R})}(\alpha_{P^0} I) \succeq \mathcal{N}(\mathcal{R}) \tag{36}$$

where

$$\pi(\mathcal{R}) = \begin{cases} \sum_{j=1}^t \mathbb{I}_{\{0\}}(i_j) & \text{if } t \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{37}$$

counts the number of f_0's in $\mathcal{R}$.

B. Some approximation results

In this subsection we present several approximation results to be used in the sequel. The results are of independent interest and establish some useful properties of the RRE and the classical Riccati operator.

The first concerns uniform convergence properties of the classical Riccati operator and is used in the sequel to obtain various tightness estimates required for establishing the MDP. The proof is provided in Appendix A.

Lemma 3.7 For every $\varepsilon > 0$, there exists $t_\varepsilon \geq N$, such that, for every $X \in \mathbb{S}_+^N$, with $X \succeq P^*$,

$$\left\| f_1^t(X) - P^* \right\| \leq \varepsilon, \quad t \geq t_\varepsilon \tag{38}$$

Note, in particular, that t_ε can be chosen independently of the initial state X.

The following result can be viewed as a corollary to Lemma 3.7 and concerns the Lipschitz continuity of finite compositions of the Riccati operator.

Lemma 3.8 For fixed $t \in \mathbb{N}$ and $i_1, \cdots, i_t \in \{0,1\}$, define the function $g : \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ by

$$g(X) = f_{i_1} \circ \cdots \circ f_{i_t}(X), \quad X \in \mathbb{S}_+^N \tag{39}$$

Then $g(\cdot)$ is Lipschitz continuous with some constant $K_g > 0$.

Also, for every $\varepsilon_2 > 0$, there exists t_{ε_2}, such that, the function $f_1^{t_{\varepsilon_2}}(\cdot)$ is Lipschitz continuous with constant $K_{f_1^{t_{\varepsilon_2}}} < \varepsilon_2$.

Proof: From (174) it follows that the function $f_1(\cdot)$ is Lipschitz continuous with constant $K_{f_1} = c_1 e^{-c_2}$, where c_1, c_2 are positive constants defined in Lemma 3.7. It is also easy to see that the affine function $f_0(\cdot)$ is Lipschitz continuous with constant $K_{f_0} = \alpha^2$, where α is the largest singular value of the matrix A.

It then follows that the function $g(\cdot)$ defined above is Lipschitz continuous, being a finite composition of Lipschitz continuous functions.

For the second assertion, choose $t_{\varepsilon_2} \in \mathbb{N}$, such that, $c_1 e^{-c_2 t_{\varepsilon_2}} < \varepsilon_2$, where $c_1, c_2 > 0$ are defined in Lemma 3.7, equation (174). It then follows from (174) that, with the above choice of t_{ε_2}, the function $f_1^{t_{\varepsilon_2}}(\cdot)$ is Lipschitz continuous with constant $K_{f_1}^{t_{\varepsilon_2}} < \varepsilon_2$. ■

The following result concerns stochastic boundedness of the random sequence $\{P_t\}_{t\in\mathbb{T}_+}$ generated by the RRE. In particular, it completes the proof of Proposition 8 in [2] by establishing triviality of $\overline{\gamma}^{\text{sb}}$ for general observable and controllable systems (in [2], Proposition 8 was proved only for systems with invertible C.)

Lemma 3.9 Assume $\left(A, Q^{1/2}\right)$ controllable, (C, A) observable. Then $\overline{\gamma}^{\text{sb}} = 0$, i.e., the sequence $\{P_t\}$ is stochastically bounded

$$\lim_{M\to\infty} \sup_{t\in\mathbb{T}_+} \mathbb{P}^{\overline{\gamma},P_0}\left(\|P_t\| > M\right) = 0 \tag{40}$$

for every $\overline{\gamma} > 0$ and initial covariance state $P_0 \in \mathbb{S}_+^N$.

The proof is provided in Appendix B and offers a new insight into the random Riccati equation. For a discussion on the consequence and significance of the result, the reader is referred to the text following Proposition 8 in [2]. We reemphasize here, that the result establishes the importance of stochastic boundedness as a metric for system stability and design as compared to various notions of moment stability. For example, as shown in [1], the critical probability for mean stability may be quite large, depending on the instability of A. However, we show that, even in the sub-mean stability regime, the system is stochastically bounded (for $\overline{\gamma} > 0$) and converges to a unique invariant distribution. Hence, our analysis offers insight into system design in the sub-mean stability regime.

The following result on limits of real number sequences will be useful later.

Proposition 3.10 For $J \in \mathbb{N}$ and $1 \le i \le J$, let $a_i : [0,1) \longmapsto [0,1]$ be functions with

$$\lim_{\overline{\gamma}\uparrow 1} -\frac{\ln a_i(\overline{\gamma})}{\ln(1-\overline{\gamma})} = -a_i^*, \qquad 1 \le i \le J \tag{41}$$

(we adopt the convention $\ln 0 = -\infty$ and a_i^* is non-negative with ∞ as a possible value. Then

$$\lim_{\overline{\gamma}\uparrow 1} -\frac{\ln\left(\sum_{i=1}^J a_i(\overline{\gamma})\right)}{\ln(1-\overline{\gamma})} = -\min_{i\in\{0,\cdots,J\}}\{a_i^*\} \tag{42}$$

4. Main Results and Discussions

We state the main results of the paper in this section whose proofs are provided in Section 7.

The following result is a first step to understanding the behavior of the family $\{\mu^{\overline{\gamma}}\}$ of invariant distributions.

Theorem 4.1 The family of invariant distributions $\left\{\mu^{\overline{\gamma}}\right\}$ converges weakly to the Dirac probability measure δ_{P^*} as $\overline{\gamma} \uparrow 1$, i.e.,

$$\lim_{\overline{\gamma}\uparrow 1} d_P\left(\mu^{\overline{\gamma}}, \delta_{P^*}\right) = 0 \tag{43}$$

We have the following convergence rate asymptotics:

For every $\varepsilon > 0$, we have

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{\ln\left(\mu^{\overline{\gamma}}\left(B_\varepsilon^C(P^*)\right)\right)}{\ln(1-\overline{\gamma})} \le -1 \tag{44}$$

We discuss the consequences of Theorem 4.1. The first assertion states that the family $\{\mu^{\overline{\gamma}}\}$ converges weakly to the Dirac measure concentrated at P^*, δ_{P^*}, as $\overline{\gamma} \uparrow 1$. This is quite intuitive, as with $\overline{\gamma} \uparrow 1$, the filtering problem

reduces to the classical Kalman filtering setup with deterministic packet arrival (no dropouts), in which case the deterministic sequence of conditional error covariances converges to the fixed point P^*. Thus, as $\overline{\gamma} \uparrow 1$, we expect the RRE sequence to behave more and more similarly to the deterministic $\overline{\gamma} = 1$ case leading to the convergence of the measures $\mu^{\overline{\gamma}}$ to δ_{P^*} as $\overline{\gamma} \uparrow 1$. However, from a technical point of view this is not obvious, as the case $\overline{\gamma} = 1$ may be a singularity. Theorem 4.1 rules out this possibility and shows that the family $\{\mu^{\overline{\gamma}}\}$ viewed as a function of $\overline{\gamma}$ is sufficiently well behaved (continuous) at $\overline{\gamma} = 1$.

An immediate consequence of Theorem 4.1 is the following:

$$\lim_{\overline{\gamma}\uparrow 1} \mu^{\overline{\gamma}}(\Gamma) = 0, \quad \forall\ \overline{\Gamma} \cap P^* = \phi \tag{45}$$

Thus, we note that, w.r.t. $\{\mu^{\overline{\gamma}}\}$, every event Γ with $P^* \notin \overline{\Gamma}$ is a rare event (see Defn. (1.2).) This is intuitively clear, because as $\overline{\gamma} \uparrow 1$, the measures $\mu^{\overline{\gamma}}$ become concentrated on arbitrarily small neighborhoods of P^*, making such an event Γ very difficult to observe.

Once the rare events are identified, the next step in the characterization of $\{\mu^{\overline{\gamma}}\}$ is to ascertain the rate at which such rare events go to zero, or the rate at which the family $\{\mu^{\overline{\gamma}}\}$ converges to δ_{P^*} as $\overline{\gamma} \uparrow 1$. The distance between the family $\{\mu^{\overline{\gamma}}\}$ and its weak limit δ_{P^*} as $\overline{\gamma} \uparrow 1$ is important for the design engineer, as it relates the loss in performance when operating at packet arrival probability $\overline{\gamma} < 1$. The answer is provided in the second assertion of Theorem 4.1, which says that, for every $\varepsilon > 0$, the probability of staying away from the ε-neighborhood of P^* decreases as $(1 - \overline{\gamma})$ when $\overline{\gamma} \uparrow 1$, i.e.[6],

$$\mu^{\overline{\gamma}} = O(1 - \overline{\gamma}), \quad \forall\ \varepsilon > 0 \tag{46}$$

Since the above holds for every $\varepsilon > 0$, the probability of any rare event vanishes at least as $(1 - \overline{\gamma})$. Clearly, the exact rate of going to zero depends on the rare event in question (for example, how 'far' it is from P^*, which becomes 'typical' as $\overline{\gamma} \uparrow 1$.

A complete characterization of the family $\{\mu^{\overline{\gamma}}\}$ requires the exact decay rate of rare events as $\overline{\gamma} \uparrow 1$, and this is achieved in the following result that establishes an MDP for the family $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \uparrow 1$.

Theorem 4.2 Recall in equation (37) the definition of $\pi(\cdot)$. The family of invariant distributions $\left\{\mu^{\overline{\gamma}}\right\}$ satisfies an MDP at scale $-\ln(1 - \overline{\gamma})$ as $\overline{\gamma} \uparrow 1$ with a good rate function $I(\cdot)$, i.e.,

$$\liminf_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}}(\mathcal{O}) \geq -\inf_{X \in \mathcal{O}} I(X), \quad \text{for every open set } \mathcal{O} \tag{47}$$

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}}(\mathcal{F}) \leq -\inf_{X \in \mathcal{F}} I(X), \quad \text{for every closed set } \mathcal{F} \tag{48}$$

where the function $I : \mathbb{S}_+^N \longmapsto \overline{\mathbb{R}}_+$ is given by:

$$I(X) = \inf_{\mathcal{R} \in \mathcal{S}^{P^*}} \pi(\mathcal{R}), \quad \forall X \in \mathbb{S}_+^N \tag{49}$$

Theorem 4.2 provides a complete understanding of the family of invariant measures $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \uparrow 1$. First, it establishes the important qualitative behavior of $\{\mu^{\overline{\gamma}}\}$, namely, that rare events decay *exactly* as power-laws of $(1 - \overline{\gamma})$ as $\overline{\gamma} \uparrow 1$. Also the exact exponent of such a power law decay depends on the particular rare event and is obtained as the solution of an associated variational problem involving the minimization of the rate function $I(\cdot)$. This is relevant for a system designer who can trade-off estimation accuracy with communication required. For

[6]For functions $h(\cdot), g(\cdot)$, the notation $h(\overline{\gamma}) = O(g(\overline{\gamma}))$ implies the existence of a constant $c > 0$, such that $h(\overline{\gamma}) \leq cg(\overline{\gamma})$ for all $\overline{\gamma} \in (0, 1)$.

example, given a tolerance $M > 0$, we may ask the question, at what operating $\overline{\gamma}$ is the probability of lying outside the M-neighborhood centered at P^* less than some $\delta > 0$. Using the notation of (12), we then have

$$\mu^{\overline{\gamma}}\left(B_M^C(P^*)\right) \sim (1-\overline{\gamma})^{\inf_{X \in B_M^C(P^*)} I(X)} \tag{50}$$

Thus by computing $\inf_{X \in B_M^C(P^*)} I(X)$, the designer obtains an estimate of the $\overline{\gamma}$ required to maintain a probability of error less than δ. Thus the estimation of probabilities of rare events reduces to solving deterministic variational problems. As shown in Section 8, several techniques can be employed to solve these variational problems efficiently. Finally, we emphasize that our analysis of reducing the problem of estimating probabilities of interest to solving variational problems efficiently is much more definitive and relevant than numerically estimating the invariant distributions. A naive numerical approach of simulating the distributions $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \uparrow 1$ becomes meaningless here as the rare events, which are of interest, become increasingly difficult to simulate as $\overline{\gamma} \uparrow 1$. One may take recourse to sophisticated simulation techniques like importance sampling (see, for example, [14]), but such approaches require characterization of the distributions in question, which is addressed in this paper.

5. Moderate Deviations (MD) for finite-dimensional distributions

In this section, we establish MD for finite dimensional distributions of the process $\{P_t\}_{t \in \mathbb{T}_+}$ as $\overline{\gamma} \uparrow 1$. We start by setting notation. Fix $n \in \mathbb{N}$ and $t_1 < \cdots < t_n \in \mathbb{T}_+$ and recall the sets $\mathcal{S}^{P_0}_{t_1,\cdots,t_n}$ and $\mathcal{S}_{t_1,\cdots,t_n}(\mathbf{X})$ for $\mathbf{X} \in \bigotimes_{i=1}^n \mathbb{S}_+^N$. As noted in Proposition 3.6, the random object $(P_{t_1}, \cdots, P_{t_n})$ takes only a finite number of values and for all $\overline{\gamma}$

$$\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1}, \cdots, P_{t_n}) \in \mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)\right) = 1 \tag{51}$$

We generalize the definition of the functional $\pi : \mathcal{S}_t^{P_0} \longmapsto \mathbb{Z}_+$ (eqn. 37) to strings in $\mathcal{S}^{P_0}_{t_1,\cdots,t_n}$ by

$$\pi\left(\overline{\mathcal{R}}\right) = \begin{cases} \sum_{j=1}^{t_n} \mathbb{I}_{\{0\}}(i_j) & \text{if } t \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

Thus $\pi(\cdot)$ counts the number of f_0's in the string $\overline{\mathcal{R}}$. Also, for $\mathbf{X} \in \bigotimes_{i=1}^n \mathbb{S}_+^N$, define

$$\ell(\mathbf{X}) = \min_{\overline{\mathcal{R}} \in \mathcal{S}^{P_0}_{t_1,\cdots,t_n}(\mathbf{X})} \pi(\overline{\mathcal{R}}) \tag{53}$$

(We adopt the convention that the minimum of an empty set is ∞.)

The following result shows that the function $\ell : \bigotimes_{i=1}^n \mathbb{S}_+^N \longmapsto \mathbb{R}_+$ is a good rate function on $\bigotimes_{i=1}^n \mathbb{S}_+^N$.

Proposition 5.1 The function $\ell : \bigotimes_{i=1}^n \mathbb{S}_+^N \longmapsto \mathbb{R}_+$ in (53) is a good rate function on $\bigotimes_{i=1}^n \mathbb{S}_+^N$.

Proof: Clearly, $\ell(\cdot) > 0$. Its level sets are compact because its effective domain $\mathcal{D}_\ell$ is finite, where

$$\mathcal{D}_\ell = \left\{ \mathbf{X} \in \bigotimes_{i=1}^n \mathbb{S}_+^N \mid \ell(\mathbf{X}) < \infty \right\} \tag{54}$$

■

We then have the following result giving the MD for the family (as $\overline{\gamma} \uparrow 1$) of finite dimensional distributions $(P_{t_1}, \cdots, P_{t_n})$. The proof is provided in Appendix C.

Theorem 5.2 Fix $n \in \mathbb{N}$ and $t_1 < \cdots < t_n \in \mathbb{T}_+$ and let $(P_{t_1}, \cdots, P_{t_n})^{\overline{\gamma}}$ be the family of of finite dimensional distributions indexed by $\overline{\gamma}$, starting from the same initial state P_0. Then for every $B \in \mathcal{B}\left(\bigotimes_{i=1}^{n} \mathbb{S}_+^N\right)$, we have

$$\lim_{\overline{\gamma}\uparrow 1} \frac{1}{\ln(1-\overline{\gamma})} \ln\left(\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1}, \cdots, P_{t_n})^{\overline{\gamma}} \in B\right)\right) = \inf_{\mathbf{X}\in B} I_{P_0}^{t_1,\cdots,t_n}(\mathbf{X}) \tag{55}$$

where

$$I_{P_0}^{t_1,\cdots,t_n}(\mathbf{X}) = \begin{cases} \ell(\mathbf{X}) & \text{if } \mathbf{X} \in \mathcal{S}_{t_1,\cdots,t_n}^{P_0} \\ \infty & \text{otherwise} \end{cases} \tag{56}$$

Remark 5.3 The MDP for finite dimensional distributions gives insight into transient system behavior. More importantly, it offers considerable insight in guessing the proper rate function governing the MDP for the family of invariant distributions. A naive approach to the MDP rate function for the family of invariant distributions is to view it as a suitable 'limit' of fixed time rate functions $I^t(\cdot)$, provided the latter converges in an appropriate sense to a limit $I(\cdot)$ as $t \to \infty$. The intuition being the weak convergence of the random sequence $\{P_t\}_{t\in\mathbb{T}_+}$ as $t \to \infty$. However, in general, the limit of fixed time rate functions may not be the rate function governing the MDP for the family of invariant rate functions. One needs to verify rigorously that the guessed rate function obtained intuitively is the actual rate function governing the required MDP. This is precisely the way (at least implicitly) we establish the MDP rate function of the family of invariant distributions in Section 6.

6. MDP FOR INVARIANT DISTRIBUTIONS

This section constitutes the key technical part of the paper. Apart from establishing the main ingredients for proving the MDP results in Section 4, the results are of independent interest and form a basis for understanding the characteristics of stationary measures resulting from *stable* iterated function systems in general. As suggested in Remark 5.3, an intuitive guess for the MDP rate function of the family $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \uparrow 1$ is $I(\cdot)$. However, apriori, it is not even obvious whether $I(\cdot)$ is lower semicontinuous to qualify as a rate function. Hence, we start by defining the lower semicontinuous regularization I_L of I and establish some of its properties in Subsection 6-A. At this point, it is not clear whether $I = I_L$ (i.e., I is lower semicontinuous) and, hence, we set to establish the MDP for the family $\{\mu^{\overline{\gamma}}\}$ with I_L as a candidate rate function. The major technical lemmas of this section are presented in Subsections 6-B,6-C, where we establish the MDP lower and upper bounds respectively w.r.t. the proposed rate function I_L. Our approach is fairly general and is not necessarily restricted to the particular filtering problem considered here.

A. A rate function

Recall: $I : \mathbb{S}_+^N \longmapsto \overline{\mathbb{R}}_+$ by

$$I(X) = \inf_{\mathcal{R}\in\mathcal{S}^{P^*}} \pi(\mathcal{R}), \quad \forall X \in \mathbb{S}_+^N \tag{57}$$

where, as usual, we adopt the convention that the infimum of an empty set is ∞. Note that, for $X \in \mathbb{S}_+^N$,

$$I(X) = \inf_{t\in\mathbb{T}_+} I_t^{P^*}(X) \tag{58}$$

and can be thought of as a natural generalization of the marginal rate functions $I_t^{P^*}(\cdot)$ for all t.

However, the function $I(\cdot)$ is not generally lower semicontinuous (as will be seen later) and hence does not qualify as a rate function. A candidate rate function for the family of invariant distributions can be the lower

semicontinuous regularization of $I(\cdot)$, defined as

$$I_L(X) = \lim_{\varepsilon \to 0} \inf_{Y \in B_\varepsilon(X)} I(Y), \quad \forall X \in \mathbb{S}_+^N \tag{59}$$

(note if $I(\cdot)$ is lower semicontinuous then $I(\cdot) = I_L(\cdot)$.)

The following proposition states some easily verifiable properties of $I_L : \mathbb{S}_+^N \longmapsto \overline{\mathbb{R}}_+$, whose proof is provided in Appendix D.

Proposition 6.1 (i) The function $I_L(\cdot)$ is a good rate function on $\mathbb{S}_+^N$.

(ii) For every $X \in \mathbb{S}_+^N$, we have

$$I_L(X) = \lim_{\varepsilon \to 0} \inf_{Y \in \overline{B_\varepsilon(X)}} I(Y) \tag{60}$$

(iii) For any non-empty set $\Gamma \in \mathcal{B}(\mathbb{S}_+^N)$ we have

$$\inf_{X \in \Gamma} I_L(X) \leq \inf_{X \in \Gamma} I(X) \tag{61}$$

In addition, if Γ is open, the reverse inequality holds and we have

$$\inf_{X \in \Gamma} I_L(X) = \inf_{X \in \Gamma} I(X) \tag{62}$$

(iv) Let $K \subset \mathbb{S}_+^N$ be a non-empty compact set. We have

$$\lim_{\varepsilon \to 0} \inf_{Y \in \overline{K_\varepsilon}} I_L(Y) = \inf_{Y \in K} I_L(Y) \tag{63}$$

B. The MDP lower bound

The following result establishes the MDP lower bound for the sequence $\{\mu^{\overline{\gamma}}\}$ of invariant distributions as $\overline{\gamma} \uparrow 1$.

Lemma 6.2 Let $\Gamma \in \mathcal{B}\left(\mathbb{S}_+^N\right)$. Then the following lower bound holds:

$$\liminf_{\overline{\gamma} \uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}}(\Gamma) \geq -\inf_{X \in \Gamma^\circ} I_L(X) \tag{64}$$

Proof: Let P_0 be an arbitrary initial state and $\left\{P_t^{\overline{\gamma},P_0}\right\}_{t \in \mathbb{T}_+}$ be the sequence generated by the RRE for $\overline{\gamma} \in (0,1)$. It was shown in Theorem 9 in [2], that, for such $\overline{\gamma}$, the sequence $\left\{P_t^{\overline{\gamma},P_0}\right\}_{t \in \mathbb{T}_+}$ converges weakly to an invariant distribution $\mu^{\overline{\gamma}}$, i.e.,

$$\liminf_{t \to \infty} \mathbb{P}^{\overline{\gamma},P_0}\left(P_t^{\overline{\gamma},P_0} \in \mathcal{O}\right) \geq \mu^{\overline{\gamma}}(\mathcal{O}), \quad \forall \text{ open set } \mathcal{O} \subset \mathbb{S}_+^N \tag{65}$$

$$\limsup_{t \to \infty} \mathbb{P}^{\overline{\gamma},P_0}\left(P_t^{\overline{\gamma},P_0} \in \mathcal{F}\right) \leq \mu^{\overline{\gamma}}(\mathcal{F}), \quad \forall \text{ closed set } \mathcal{F} \subset \mathbb{S}_+^N \tag{66}$$

Now consider the measurable set $\Gamma \in \mathcal{B}\left(\mathbb{S}_+^N\right)$ and let $X \in \Gamma^\circ \cap \mathcal{D}_I$, where $\mathcal{D}_I$ is the effective domain of $I(\cdot)$. Then, there exists $\varepsilon > 0$, sufficiently small, such that the closed ball $\overline{B}_\varepsilon(X) \in \Gamma$. From (66) it then follows

$$\mu^{\overline{\gamma}}(\Gamma) \geq \mu^{\overline{\gamma}}\left(\overline{B}_\varepsilon(X)\right) \geq \limsup_{t \to \infty} \mathbb{P}^{\overline{\gamma},P_0}\left(P_t^{\overline{\gamma},P_0} \in \mathcal{F}\right) \tag{67}$$

We now set to estimate the R.H.S. of (67).

To this end, recall the nonempty set $\mathcal{S}^{P^*}(X)$ of all strings of finite (but arbitrary) length with initial state P^* and numerical value X. For some $t_0 \in \mathbb{T}_+$ and $i_1, \cdots, i_{t_0} \in \{0,1\}$, let the string $\mathcal{R} = f_{i_1} \circ \cdots \circ f_{i_{t_0}}(P^*) \in \mathcal{S}^{P^*}(X)$.

Define the function $g : \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ by

$$g(Y) = f_{i_1} \circ \cdots \circ f_{i_{t_0}}(Y), \quad \forall Y \in \mathbb{S}_+^N \tag{68}$$

The function $g(\cdot)$ is continuous (being the composition of continuous functions) and hence there exists $\varepsilon_1 > 0$, such that

$$\|g(Y) - g(P^*)\| \leq \varepsilon, \quad \forall Y \in \overline{B}_{\varepsilon_1}(P^*) \tag{69}$$

Also, by Lemma 3.7, there exists $t_{\varepsilon_1} \in \mathbb{T}_+$, such that

$$\left\| f_1^t(Y) - P^* \right\| \leq \varepsilon_1, \quad \forall t \geq t_{\varepsilon_1}, \quad Y \in \mathbb{S}_+^N \tag{70}$$

It then follows from eqns. (69,70), that, for any $t \in \mathbb{T}_+$, such that $t \geq t_0 + t_{\varepsilon_1}$ and any string $\mathcal{R}_1 \in \mathcal{S}_t^{P_0}$ of the form

$$\mathcal{R}_1 = \left(f_{i_1}, \cdots, f_{i_{t_0}}, f_1^{t_{\varepsilon_1}}, f_{j_1} \cdots, f_{j_{t-t_0-t_{\varepsilon_1}}}, P_0 \right) \tag{71}$$

where $j_1, \cdots, j_{t-t_0-t_{\varepsilon_1}} \in \{0, 1\}$, we have

$$\mathcal{N}(\mathcal{R}_1) \in \overline{B}_{\varepsilon}(X) \tag{72}$$

Indeed

$$\begin{aligned} \|\mathcal{N}(\mathcal{R}_1) - X\| &= \|\mathcal{N}(\mathcal{R}_1) - \mathcal{N}(\mathcal{R})\| \\ &= \left\| g\left(f_1^{t_{\varepsilon_1}} \left(f_{j_1} \cdots \circ f_{j_{t-t_0-t_{\varepsilon_1}}}(P_0) \right) \right) - g(P^*) \right\| \\ &\leq \varepsilon \end{aligned} \tag{73}$$

where the last step follows from the fact, that,

$$\left\| f_1^{t_{\varepsilon_1}} \left(f_{j_1} \circ \cdots \circ f_{j_{t-t_0-t_{\varepsilon_1}}}(P_0) \right) - P^* \right\| \leq \varepsilon_1 \tag{74}$$

by (70).

Now for $\mathcal{R}$ (defined above), $t \geq t_0 + t_{\varepsilon_1}$, consider the set of strings

$$\mathcal{R}_t = \left\{ \left(f_{i_1}, \cdots, f_{i_{t_0}}, f_1^{t_{\varepsilon_1}}, f_{j_1} \cdots, f_{j_{t-t_0-t_{\varepsilon_1}}}, P_0 \right) \middle| \, j_1, \cdots, j_{t-t_0-t_{\varepsilon_1}} \in \{0, 1\} \right\} \tag{75}$$

(in other words, the indices $j_1, \cdots, j_{t-t_0-t_{\varepsilon_1}}$ can be arbitrary.)

It then follows from (72) above,

$$\mathcal{N}(\mathcal{R}_2) \in \overline{B}_{\varepsilon}(X), \quad \forall \mathcal{R}_2 \in \mathcal{R}_t \tag{76}$$

From the iterative construction of the sequence $\{P_t^{\overline{\gamma}, P_0}\}_{t \in \mathbb{T}_+}$ it is then obvious, for $t \geq t_0 + t_{\varepsilon_1}$,

$$\begin{aligned} \mathbb{P}^{\overline{\gamma}, P_0}\left(P_t^{\overline{\gamma}, P_0} \in \overline{B}_{\varepsilon}(X) \right) &\geq \mathbb{P}^{\overline{\gamma}, P_0}\left(P_t^{\overline{\gamma}, P_0} \in \mathcal{N}(\mathcal{R}_t) \right) \\ &= \sum_{j_1, \cdots, j_{t-t_0+t_{\varepsilon_1}} \in \{0,1\}} \left[\left(\prod_{k=1}^{t_0} (1-\overline{\gamma})^{1-i_k} \overline{\gamma}^{i_k} \right) \overline{\gamma}^{t_{\varepsilon_1}} \left(\prod_{k=1}^{t-t_0-t_{\varepsilon_1}} (1-\overline{\gamma})^{1-j_k} \overline{\gamma}^{j_k} \right) \right] \\ &= (1-\overline{\gamma})^{\pi(\mathcal{R})} \overline{\gamma}^{t_0 - \pi(\mathcal{R})} \overline{\gamma}^{t_{\varepsilon_1}} \\ &= (1-\overline{\gamma})^{\pi(\mathcal{R})} \overline{\gamma}^{t_0 + t_{\varepsilon_1} - \pi(\mathcal{R})} \end{aligned} \tag{77}$$

From eqns. (67,77) we have

$$\mu^{\overline{\gamma}}(\Gamma) \geq (1-\overline{\gamma})^{\pi(\mathcal{R})}\overline{\gamma}^{t_0+t_{\varepsilon_1}-\pi(\mathcal{R})} \tag{78}$$

Noting that $\ln(1-\overline{\gamma}) < 0$ and passing to the limit in the above we have

$$\liminf_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\Gamma) \geq -\pi(\mathcal{R}) \tag{79}$$

Since the above holds for all $\mathcal{R} \in \mathcal{S}^{P^*}(X)$, we have

$$\liminf_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\Gamma) \geq \sup_{\mathcal{R}\in\mathcal{S}^{P^*}(X)}(-\pi(\mathcal{R})) = -\inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)}(\pi(\mathcal{R})) = -I(X) \tag{80}$$

The above holds for all $X \in \Gamma^\circ \cap \mathcal{D}_I$ and hence

$$\liminf_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\Gamma) \geq -\inf_{X\in\Gamma^\circ\cap\mathcal{D}_I} I(X) = -\inf_{X\in\Gamma^\circ} I(X) \tag{81}$$

where the last step follows from the fact that, for $X \notin \mathcal{D}_I$, $I(X) = \infty$. To establish the Lemma, it suffices to show the R.H.S. of (81) satisfies

$$-\inf_{X\in\Gamma^\circ} I(X) = -\inf_{X\in\Gamma^\circ} I_L(X) \tag{82}$$

which follows from Proposition 6.1 Assertion (iii), since Γ° is an open set.

■

We extract the following result for later use, which follows from the arguments in the proof of Lemma 6-B culminating to (78).

Corollary 6.3 Let $\mathcal{O} \subset \mathbb{S}_+^N$ be an open set and $\mathcal{R} \subset \mathcal{S}^{P^*}$ be a string, such that, $\mathcal{N}(\mathcal{R}) \in \mathcal{O}$. Then, there exists a positive integer $t_{\mathcal{R},\mathcal{O}}$ (depending on $\mathcal{R}$ and $\mathcal{O}$), such that,

$$\mu^{\overline{\gamma}}(\mathcal{O}) \geq (1-\overline{\gamma})^{\pi(\mathcal{R})}\overline{\gamma}^{t_{\mathcal{R},\mathcal{O}}-\pi(\mathcal{R})}, \quad \forall\overline{\gamma} \in [0,1] \tag{83}$$

C. The MDP upper bound

In this subsection, we establish the MDP upper bound for the family of invariant measures as $\overline{\gamma} \uparrow 1$. The proof is carried out in essentially three stages. First, we establish the upper bound for compact sets and this is done in Lemma 6.7. Then, in Lemma 6.8 we prove a tightness result on the family of invariant distributions and finally establish the MDP upper bound for closed sets in Lemma 6.9.

We start with the following result on topological properties of strings. We need the following definition.

Definition 6.4 (Truncated String) Let the string $\mathcal{R}$ be given by

$$\mathcal{R} = (f_{i_1}, \cdots, f_{i_t}, P_0) \tag{84}$$

where $t \in \mathbb{T}_+$, $i_1, \cdots, i_t \in \{0,1\}$ and $P_0 \in \mathbb{S}_+^N$. Then for $s \leq t$, the truncated string $\mathcal{R}^s$ of length s is given by

$$\mathcal{R}^s = (f_{i_1}, \cdots, f_{i_s}, P_0) \tag{85}$$

Lemma 6.5 Let $\mathcal{F} \in \mathbb{S}_+^N$ be a closed set. Define the set of strings $\mathcal{U} \subset \mathcal{S}^{P^*}$ by

$$\mathcal{U}(\mathcal{F}) = \left\{\mathcal{R} \in \mathcal{S}^{P^*} \mid \mathcal{N}(\mathcal{R}) \in \mathcal{F}\right\} \tag{86}$$

and let

$$\ell(\mathcal{F}) = \inf_{\mathcal{R}\in\mathcal{U}(\mathcal{F})} \pi(\mathcal{R}) \tag{87}$$

(we adopt the convention that the infimum of an empty set is ∞.) Then, if $\ell(\mathcal{F}) < \infty$, there exists $t_{\mathcal{F}} \in \mathbb{T}_+$ sufficiently large, such that, for all $\mathcal{R} \in \mathcal{U}(\mathcal{F})$ with $\text{len}(\mathcal{R}) \geq t_{\mathcal{F}}$, we have

$$\pi(\mathcal{R}^{t_{\mathcal{F}}}) \geq \ell(\mathcal{F}) \tag{88}$$

For a proof see Appendix D. We present the following remark.

Remark 6.6 (i) It follows from the definitions that

$$\mathcal{U}(\mathcal{F}) = \cup_{X\in\mathcal{F}} \mathcal{S}^{P^*}(X) \tag{89}$$

and hence

$$\ell(\mathcal{F}) = \inf_{X\in\mathcal{F}} \inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)} \pi(\mathcal{R}) = \inf_{X\in\mathcal{F}} I(X) \tag{90}$$

(ii) If $\ell(\mathcal{F}) < \infty$, i.e., the set $\mathcal{U}(\mathcal{F})$ is non-empty, the infimum in (87) is attained, i.e., there exists $\mathcal{R}^* \in \mathcal{U}(\mathcal{F})$, such that,

$$\pi(\mathcal{R}^*) = \ell(\mathcal{F}) \tag{91}$$

This follows from the fact that the function $\pi(\cdot)$ takes only a countable number of values. Similarly, in the case $\ell(\mathcal{F}) < \infty$, if we define $X^* = \mathcal{N}(\mathcal{R}^*)$, then

$$I(X^*) = \inf_{X\in\mathcal{F}} I(X) \tag{92}$$

We now prove the MDP upper bound for the family $\{\mu^{\overline{\gamma}}\}$ as $\overline{\gamma} \uparrow 1$ over compact sets.

Lemma 6.7 Let $K \in \mathcal{B}(\mathbb{S}_+^N)$ be a compact set. Then the following upper bound holds:

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}}(K) \leq -\inf_{X\in K} I_L(X) \tag{93}$$

Proof: For every $\varepsilon > 0$, define the ε-closure $\overline{K_\varepsilon}$ and the ε-neighborhood K_ε of K by

$$\overline{K_\varepsilon} = \left\{ X \in \mathbb{S}_+^N \mid \inf_{Y\in K} \|X - Y\| \leq \varepsilon \right\} \tag{94}$$

$$K_\varepsilon = \left\{ X \in \mathbb{S}_+^N \mid \inf_{Y\in K} \|X - Y\| < \varepsilon \right\} \tag{95}$$

Since K_ε is open, we have by the weak convergence of the sequence $\{P_t^{\overline{\gamma},P^*}\}_{t\in\mathbb{T}_+}$ to $\mu^{\overline{\gamma}}$

$$\liminf_{t\to\infty} \mathbb{P}^{\overline{\gamma},P^*} \left(P_t^{\overline{\gamma},P^*} \in K_\varepsilon \right) \geq \mu^{\overline{\gamma}}(K_\varepsilon) \tag{96}$$

which in turn implies

$$\liminf_{t\to\infty} \mathbb{P}^{\overline{\gamma},P^*} \left(P_t^{\overline{\gamma},P^*} \in \overline{K_\varepsilon} \right) \geq \mu^{\overline{\gamma}}(K) \tag{97}$$

We now estimate the probabilities on the L.H.S. of (97). Since $\overline{K_\varepsilon}$ is closed, the results of Lemma 6.5 apply and recall the objects $\mathcal{U}(\cdot)$ and $\ell(\cdot)$ defined for any closed set $\mathcal{F}$ as:

$$\mathcal{U}(\mathcal{F}) = \left\{ \mathcal{R} \in \mathcal{S}^{P^*} \mid \mathcal{N}(\mathcal{R}) \in \mathcal{F} \right\} \tag{98}$$

$$\ell(\mathcal{F}) = \inf_{\mathcal{R}\in\mathcal{U}(\mathcal{F})} \pi(\mathcal{R}) \tag{99}$$

with the convention that $\ell(\mathcal{F}) = \infty$ if $\mathcal{U}(\mathcal{F})$ is empty. Also, for every $t \in \mathbb{T}_+$ and closed set $\mathcal{F}$ define the sets:

$$\mathcal{U}^t(\mathcal{F}) = \mathcal{U}(\mathcal{F}) \cap \mathcal{S}_t^{P^*} \tag{100}$$

To establish the Lemma we may consider two cases, as to whether $\ell(K) < \infty$ (i.e, $\mathcal{K}$ is non-empty) or not. We first consider the non-trivial case $\ell(K) < \infty$.

To this end, consider fixed $\varepsilon > 0$ and from Proposition 3.6 Assertion (ii) it is easy to see that

$$\mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in \overline{K_\varepsilon}\right) = \mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in \mathcal{N}\left(\mathcal{U}^t(\overline{K_\varepsilon})\right)\right) = 1 \tag{101}$$

Since $K \subset \overline{K_\varepsilon}$ and $\ell(K) < \infty$, we have $\ell(\overline{K_\varepsilon}) < \infty$. The fact that $\overline{K_\varepsilon}$ is closed and Lemma 6.5 imply there exists $t_{\overline{K_\varepsilon}} \in \mathbb{T}_+$, such that, for every string $\mathcal{R} \in \mathcal{U}(\overline{K_\varepsilon})$ with len$(\mathcal{R}) \geq t_{\overline{K_\varepsilon}}$, we have $\pi\left(\mathcal{R}^{t_{\overline{K_\varepsilon}}}\right) \geq \ell(K_\varepsilon)$. In other words, we have for all $t \geq t_{\overline{K_\varepsilon}}$,

$$\pi\left(\mathcal{R}^{t_{\overline{K_\varepsilon}}}\right) \geq \ell(K_\varepsilon), \quad \forall \mathcal{R} \in \mathcal{U}^t(\overline{K_\varepsilon}) \tag{102}$$

Now consider $t \geq t_{\overline{K_\varepsilon}}$ and define the set of strings $\mathcal{J}_t^{P^*}$ by

$$\mathcal{J}_t^{P^*} = \left\{ \mathcal{R} \in \mathcal{S}_t^{P^*} \,\middle|\, \pi\left(\mathcal{R}^{t_{\overline{K_\varepsilon}}}\right) \geq \ell(K_\varepsilon) \right\} \tag{103}$$

In other words, the set $\mathcal{J}_t^{P^*}$ consists of all strings $\mathcal{R}$ of length t, such that there are at least $\ell(\overline{K_\varepsilon})$ occurrences of f_0's in the truncated string $\mathcal{R}^{t_{\overline{K_\varepsilon}}}$. The following inclusion is then obvious for $t \geq t_{\overline{K_\varepsilon}}$:

$$\mathcal{U}^t(\overline{K_\varepsilon}) \subset \mathcal{J}_t^{P^*} \subset \mathcal{S}_t^{P^*} \tag{104}$$

By the Markovian dynamics of the RRE, it is clear, that for $t \geq t_{\overline{K_\varepsilon}}$

$$\mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in \mathcal{N}(\mathcal{J}_t^{P^*})\right) = \sum_{\mathcal{R}\in\mathcal{J}_t^{P^*}} (1-\overline{\gamma})^{\pi(\mathcal{R})}\overline{\gamma}^{t-\pi(\mathcal{R})} \leq \binom{t_{\overline{K_\varepsilon}}}{\ell(\overline{K_\varepsilon})}(1-\overline{\gamma})^{\ell(\overline{K_\varepsilon})} \tag{105}$$

We then have from eqns. (101,104)

$$\mu^{\overline{\gamma}}(K) \leq \liminf_{t\to\infty} \mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in \mathcal{N}(\mathcal{U}_t(\overline{K_\varepsilon}))\right) \leq \liminf_{t\to\infty} \mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in \mathcal{N}(\mathcal{J}_t^{P^*})\right) \leq \binom{t_{\overline{K_\varepsilon}}}{\ell(\overline{K_\varepsilon})}(1-\overline{\gamma})^{\ell(\overline{K_\varepsilon})} \tag{106}$$

Taking the log-limits on both sides and noting that $\ln(1-\overline{\gamma})$ is negative, $t_{\overline{K_\varepsilon}}$ is independent of $\overline{\gamma}$ we have

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(K) \leq -\ln\left(\binom{t_{\overline{K_\varepsilon}}}{\ell(\overline{K_\varepsilon})}\right)\lim_{\overline{\gamma}\uparrow 1}\frac{1}{\ln(1-\overline{\gamma})} - \lim_{\overline{\gamma}\uparrow 1}\ell(\overline{K_\varepsilon}) = -\ell(\overline{K_\varepsilon}) \tag{107}$$

Taking the limit on both sides as $\varepsilon \to 0$ we have

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(K) \leq -\lim_{\varepsilon\to 0}\ell(\overline{K_\varepsilon}) \tag{108}$$

From Lemma 6.1 (Assertion (iii)) we have for all $\varepsilon > 0$

$$\ell(\overline{K_\varepsilon}) = \inf_{Y\in\overline{K_\varepsilon}} I(Y) \geq \inf_{Y\in\overline{K_\varepsilon}} I_L(Y) \tag{109}$$

Taking the limit and using Lemma 6.1 (Assertion (iv)) we have

$$\lim_{\varepsilon\to 0}\ell(\overline{K_\varepsilon}) \geq \lim_{\varepsilon\to 0}\inf_{Y\in\overline{K_\varepsilon}} I_L(Y) = \inf_{Y\in K} I_L(Y) \tag{110}$$

The Lemma then follows from eqns. (108,110). ■

The following tightness result enables us to extend the upper bound from compact sets to arbitrary closed sets (see Appendix D for a proof.)

Lemma 6.8 The family of invariant distributions $\left\{\mu^{\overline{\gamma}}\right\}$ satisfies the following tightness property: For every $a>0$, there exists a compact set $K_a\subset\mathbb{S}_+^N$, such that

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\mu^{\overline{\gamma}}(K_a^C) \leq -a \tag{111}$$

We now complete the proof of the MDP upper bound for arbitrary closed sets by the upper bound for compact sets and the tightness estimate obtained in Lemma 6.8. This has parallels with the theory of large deviations, where one establishes the LDP upper bound first for compact sets. The tightness analogue of Lemma 6.8 here is called *exponential tightness* in the context of LDP. It can be shown that the LDP upper bound for closed sets follows from that of compact sets and exponential tightness, see, for example, [15]. Here, although we are concerned with a MDP, the proof philosophy is related, i.e., we first establish the MDP upper bound for compact sets and then use the tightness estimate of Lemma 6.8 to extend it to arbitrary closed sets.

Lemma 6.9 Let $\mathcal{F}\in\mathcal{B}(\mathbb{S}_+^N)$ be a closed set. Then the following upper bound holds:

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\mathcal{F}) \leq -\inf_{X\in\mathcal{F}} I_L(X) \tag{112}$$

Proof: Let $a>0$ be an arbitrary positive number. By the tightness estimate in Lemma 6.8, there exists a compact set $K_a\subset\mathbb{S}_+^N$, such that,

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\mu^{\overline{\gamma}}(K_a^C) \leq -a \tag{113}$$

The set $\mathcal{F}\cap K_a$ is compact, being the intersection of a closed and a compact set, and hence the MDP upper bound holds from Lemma 6.7, i.e., we have

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\mathcal{F}\cap K_a) \leq -\inf_{X\in\mathcal{F}\cap K_a} I_L(X) \tag{114}$$

To estimate the probabilities $\mu^{\overline{\gamma}}(\mathcal{F})$, we use the decomposition:

$$\mu^{\overline{\gamma}}(\mathcal{F}) = \mu^{\overline{\gamma}}(\mathcal{F}\cap K_a) + \mu^{\overline{\gamma}}\left(\mathcal{F}\cap K_a^C\right) \leq \mu^{\overline{\gamma}}(\mathcal{F}\cap K_a) + \mu^{\overline{\gamma}}\left(K_a^C\right) \tag{115}$$

By Lemma 3.10 we then have

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\mathcal{F}) \leq \max\left(\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}\left(\mathcal{F}\cap K_a^C\right), \limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}\left(K_a^C\right)\right) \tag{116}$$

From eqns. (94,95) we then have

$$\begin{aligned}\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\mathcal{F}) &\leq \max\left(-\inf_{X\in\mathcal{F}\cap K_a} I_L(X), -a\right)\\ &\leq \max\left(-\inf_{X\in\mathcal{F}} I_L(X), -a\right)\\ &= -\min\left(\inf_{X\in\mathcal{F}} I_L(X), a\right) \end{aligned} \tag{117}$$

Since (117) holds for all $a \in \mathbb{R}_+$, passing to the limit as $a \to \infty$ on both sides we obtain

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\ln\mu^{\overline{\gamma}}(\mathcal{F}) \leq -\inf_{X\in\mathcal{F}} I_L(X) \tag{118}$$

This establishes the MDP upper bound for arbitrary closed sets. ■

7. Proofs of Theorems

Theorem 4.2: The MDP lower and upper bounds obtained in Lemma 6.2 and Lemma 6.9 respectively show that the family $\left\{\mu^{\overline{\gamma}}\right\}$ satisfies an MDP at scale $-\ln(1-\overline{\gamma})$ as $\overline{\gamma}\uparrow 1$ with a good rate function $I_L(\cdot)$. To complete the proof of Theorem 4.2 it suffices to show that $I(\cdot) = I_L(\cdot)$, i.e., the function $I(\cdot)$ is lower semicontinuous. We now show that

$$I(X) = I_L(X), \quad \forall X \in \mathbb{S}_+^N \tag{119}$$

Clearly, if $I_L(X) = \infty$, the claim in (119) follows from (212). We thus consider the case $I_L(X) < \infty$. Since

$$I_L(X) = \lim_{\varepsilon\to 0} \inf_{Y\in B_\varepsilon(X)} I(X) \tag{120}$$

and the integer-valued quantity $\inf_{Y\in B_\varepsilon(X)} I(X)$ is non-decreasing w.r.t. ε, there exists $\varepsilon_0 > 0$, such that

$$\inf_{Y\in B_\varepsilon(X)} I(X) = I_L(X), \quad \forall \varepsilon \leq \varepsilon_0 \tag{121}$$

The infimum above is achieved for every $\varepsilon > 0$, and we conclude that there exists a sequence $\{X_n\}_{n\in\mathbb{N}}$, such that

$$X_n \in \overline{B_{\varepsilon_0}}(X), \quad \lim_{n\to\infty} X_n = X, \quad I(X_n) = I_L(X) \tag{122}$$

Recall the set of strings

$$\mathcal{U}(\overline{B_{\varepsilon_0}}(X)) = \{\mathcal{R} \in \mathcal{S}^{P^*} \mid \mathcal{N}(\mathcal{R}) \in \overline{B_{\varepsilon_0}}(X)\} \tag{123}$$

We then have

$$\ell(\overline{B_{\varepsilon_0}}(X)) = \inf_{Y\in\overline{B_{\varepsilon_0}}(X)} I(X) = I_L(X) \tag{124}$$

Since $\overline{B_{\varepsilon_0}}(X)$ is closed, by Lemma 6.5, there exists $t_0 \in \mathbb{T}_+$, such that, for $\mathcal{R} \in \mathcal{U}(\overline{B}_{\varepsilon_0}(X))$ with $\text{len}(\mathcal{R}) \geq t_0$, we have

$$\pi\left(\mathcal{R}^{t_0}\right) \geq \ell(\overline{B}_{\varepsilon_0}) = I_L(X) \tag{125}$$

By the existence of $\{X_n\}$, there exists a sequence $\{\mathcal{R}_n\}$ of strings in $\mathcal{U}(\overline{B}_{\varepsilon_0})$, such that

$$\mathcal{N}(\mathcal{R}_n) = X_n, \quad \pi(\mathcal{R}_n) = I_L(X) \tag{126}$$

Note that, without loss of generality, we can assume that $\text{len}(\mathcal{R}_n) = t_0$ for all n. Indeed, if $\text{len}(\mathcal{R}_n) < t_0$, we can modify $\mathcal{R}_n$ by appending the requisite number of f_1's at the right end, still satisfying (126). On the other hand, if $\text{len}(\mathcal{R}_n) > t_0$, we note that $\mathcal{R}_n$ must be of the form

$$\mathcal{R}_n = \left(f_{i_1}, \cdots, f_{i_{t_0}}, f_1^{\text{len}(\mathcal{R}_n)-t_0}, P^*\right) \tag{127}$$

where the truncated string

$$\mathcal{R}_n^{t_0} = f_{i_1} \circ \cdots \circ f_{i_{t_0}}(P^*) \tag{128}$$

satisfies

$$\mathcal{N}\left(\mathcal{R}_n^{t_0}\right) = X_n, \quad \pi\left(\mathcal{R}_n^{t_0}\right) = I_L(X) \tag{129}$$

The second inequality in (129) follows from (125), whereas (128) follows from the fact that $\pi(\mathcal{R}_n) = I_L(X)$ implying

$$\pi(\mathcal{R}_n) = \pi(\mathcal{R}_n^{t_0}) \tag{130}$$

Thus the right end of $\mathcal{R}_n$ does not contain any f_0, explaining the form in (128). The key conclusion of the above discussion is that, if $\text{len}(\mathcal{R}_n) > t_0$, we may consider the truncated string $\mathcal{R}_n^{t_0}$ instead, which also satisfies (126).

We thus assume that the sequence $\{\mathcal{R}_n\}$ with the properties in (126) satisfy:

$$\text{len}(\mathcal{R}_n) = t_0, \quad \forall n \tag{131}$$

The number of distinct strings in the sequence $\{\mathcal{R}_n\}$ is at most 2^{t_0} (in fact, lesser than that, because of the constraint $\pi(\mathcal{R}_n) = I_L(X)$) and, hence, at least one pattern is repeated infinitely often in the sequence $\{\mathcal{R}_n\}$, i.e., there exists a string $\mathcal{R}^*$, such that,

$$\text{len}(\mathcal{R}^*) = t_0, \quad \pi\left(\mathcal{R}^*\right) = I_L(X) \tag{132}$$

and a subsequence $\{\mathcal{R}_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mathcal{R}_n\}$, such that,

$$\mathcal{R}_{n_k} = \mathcal{R}^*, \quad \forall k \in \mathbb{N} \tag{133}$$

The corresponding subsequence $\{X_{n_k}\}$ of numerical values then satisfy

$$X_{n_k} = \mathcal{N}(\mathcal{R}_{n_k}) = \mathcal{N}(\mathcal{R}^*), \quad \forall k \in \mathbb{N} \tag{134}$$

and hence

$$X = \lim_{k\to\infty} X_{n_k} = \mathcal{N}(\mathcal{R}^*) \tag{135}$$

Thus the string $\mathcal{R}^* \in \mathcal{S}^{P^*}(X)$ and hence

$$I(X) = \inf_{\mathcal{R}\in\mathcal{S}^{P^*}(X)} \pi(\mathcal{R}) \leq \pi\left(\mathcal{R}^*\right) = I_L(X) \tag{136}$$

The other inequality $I_L(X) \leq I(X)$ is obvious and hence we conclude from (136) that

$$I_L(X) = I(X) \tag{137}$$

This completes the proof of Theorem 4.2. ■

Theorem 4.1: Recall

$$d_P\left(\mu^{\overline{\gamma}}, \delta_{P^*}\right) = \inf\left\{\varepsilon > 0 \mid \delta_{P^*}(\mathcal{F}) \leq \mu^{\overline{\gamma}}(\mathcal{F}_\varepsilon) + \varepsilon, \quad \forall \text{ closed set } \mathcal{F}\right\} \tag{138}$$

Define the class of sets

$$\mathcal{C} = \{\mathcal{F} \mid \mathcal{F} \text{ is closed and } P^* \in \mathcal{F}\} \tag{139}$$

Then the following equivalence is straight forward:

$$d_P\left(\mu^{\overline{\gamma}}, \delta_{P^*}\right) = \inf\left\{\varepsilon > 0 \mid \mu^{\overline{\gamma}}(\mathcal{F}_\varepsilon) + \varepsilon \geq 1, \quad \forall\, \mathcal{F} \in \mathcal{C}\right\} \tag{140}$$

Now consider $0 < \varepsilon < 1$, small enough. Then there exists $\varepsilon_0 > 0$, such that for every $\mathcal{F} \in \mathcal{C}$, we have $B_{\varepsilon_0}(P^*) \subset \mathcal{F}_\varepsilon$. (Note that the constant ε_0 can be chosen independently of $\mathcal{F}$, but depends on ε.) The string $\mathcal{R} = P^*$ belongs to $B_{\varepsilon_0}(P^*)$ and hence by Corollary 6.3, there exists an integer $t_0 > 0$, such that,

$$\mu^{\overline{\gamma}}\left(B_{\varepsilon_0}(P^*)\right) \geq (1-\overline{\gamma})^{\pi(\mathcal{R})}\overline{\gamma}^{t_0-\pi(\mathcal{R})} = \overline{\gamma}^{t_0} \tag{141}$$

Thus for all $\mathcal{F} \in \mathcal{C}$

$$\mu^{\overline{\gamma}}\left(\mathcal{F}_\varepsilon\right)) \geq \mu^{\overline{\gamma}}\left(B_{\varepsilon_0}(P^*)\right) \geq \overline{\gamma}^{t_0} \tag{142}$$

Then for $\overline{\gamma} \geq (1-\varepsilon)^{1/t_0}$ we have for all $\mathcal{F}$

$$\mathcal{F}_\varepsilon + \varepsilon \geq \overline{\gamma}^{t_0} + \varepsilon \geq 1 \tag{143}$$

It then follows from (140) that

$$d_P\left(\mu^{\overline{\gamma}}, \delta_{P^*}\right) \leq \varepsilon, \quad \overline{\gamma} \geq (1-\varepsilon)^{1/t_0} \tag{144}$$

Hence

$$\limsup_{\overline{\gamma}\uparrow 1} d_P\left(\mu^{\overline{\gamma}}, \delta_{P^*}\right) \leq \varepsilon \tag{145}$$

Since $\varepsilon > 0$ is arbitrary, by passing to the limit as $\varepsilon \to 0$, we have the weak convergence

$$\lim_{\overline{\gamma}\uparrow 1} d_P\left(\mu^{\overline{\gamma}}, \delta_{P^*}\right) = 0 \tag{146}$$

For the second assertion we note that for $\varepsilon > 0$, the closed set $B_\varepsilon^C(P^*)$ does not contain P^*. Hence $\ell(B_\varepsilon^C(P^*)) \geq 1$. The claim in (44) then follows from the MDP upper bound for closed sets (Lemma 6.9.) ■

8. Computations with the rate function

A complete characterization of probabilities under the invariant distributions are obtained in Theorem 4.2, which shows that the probability of 'rare events'[7] decays as powers of $\ln(1-\overline{\gamma})$ as $\overline{\gamma} \uparrow 1$. The best exponent of this power law decay is characterized by the MDP rate function $I(\cdot)$. As Theorem 4.2 shows, the best decay exponent of such a rare event can be computed as the infimum of the $I(\cdot)$ over that set. This reduces the problem of estimating probabilities under the invariant distributions to solutions of a related variational problem, namely, minimizing $I(\cdot)$ over Borel sets in $\mathbb{S}_+^N$. From a numerical analysis point of view, one may simulate the function $I(\cdot)$ and use a look-up table to numerically solve the associated variational problems. In this section, we show that, for a class of events of interest, the variational problem of computing the best decay exponent can be simplified to a great extent.

[7]The term rare event in this context refers to a Borel set (event) bounded away from P^*.

In particular, we are interested in estimating probabilities of the form $\mu^{\overline{\gamma}}(B_\varepsilon^C(P^*))$ for every $\varepsilon > 0$. Theorem 4.1 shows that such probabilities decay to zero at least as $(1-\overline{\gamma})$, i.e., for every $\varepsilon > 0$, we have

$$\mu^{\overline{\gamma}}(B_\varepsilon^C(P^*)) \equiv O(1-\overline{\gamma}) \tag{147}$$

However, this upper bound becomes loose as ε increases and the best exponent of decay, i.e., the best power of $(1-\overline{\gamma})$ in (147) can be computed precisely by solving the variational problem in Theorem 4.2. The main result of this section shows that, for a particular class of systems, this computation can be extremely simplified and for general systems we present bounds on the decay exponent that are simple to obtain in contrast to solving the full-fledged variational problem. For a system designer, the probabilities of the form in (147) are of interest and a technique to obtain them efficiently is of much use. We start by setting some notation.

Define the function $\iota : \mathbb{R}_+ \longmapsto \mathbb{T}_+$ by

$$\iota(M) = \inf\left\{k \in \mathbb{T}_+ \mid \left\|f_0^k(P^*) - P^*\right\| \geq M\right\} \tag{148}$$

Also, define $\iota_+ : \mathbb{R}_+ \longmapsto \mathbb{T}_+$ by

$$\iota_+(M) = \inf\left\{k \in \mathbb{T}_+ \mid \left\|f_0^k(P^*) - P^*\right\| > M\right\} \tag{149}$$

We note that $\iota(\cdot)$ is a non-decreasing right continuous function, and we have for all $M > 0$

$$\iota_+(M) \leq \iota(M) + 1, \qquad \lim_{U<M:U\to M} \iota(U) = \iota_+(M) \tag{150}$$

Also recall:

$$B_M^C(P^*) = \{X \in \mathbb{S}_+^N \mid \|X - P^*\| \geq M\} \tag{151}$$

$$\overline{B}_M^C(P^*) = \{X \in \mathbb{S}_+^N \mid \|X - P^*\| > M\} \tag{152}$$

Definition 8.1 (Class $\mathcal{A}$ systems) : Let (A, Q, C, R) be a system satisfying the assumption **(E.1)**. Then the system is called a class $\mathcal{A}$ system if

$$\mathcal{S}^- \supset \{X \in \mathbb{S}_+^N \mid X \succeq f_0(P^*)\} \tag{153}$$

where S^- is defined in (169).

We then have the following MDP asymptotics for class $\mathcal{A}$ systems (see Appendix E for a proof):

Lemma 8.2 Let (A, Q, C, R) be a class $\mathcal{A}$ system. Then we have for all $M > 0$

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}}\left(B_M^C(P^*)\right) \leq -\iota(M) \tag{154}$$

and

$$\liminf_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}}\left(\overline{B}_M^C(P^*)\right) \geq -\iota_+(M) \tag{155}$$

Remark 8.3 Lemma 8.2 shows that for class $\mathcal{A}$ systems the variational problem of computing the best decay exponent can be simplified to a great extent. In particular, rather than generating the set of all possible strings $\mathcal{S}^{P^*}$ (that grows exponentially with the length of the strings) one can systematically look into strings of the form $f_0^k(P^*),\ k \in \mathbb{T}_+$ and obtain the decay exponent. The next natural question is whether there exists a suitable characterization of class $\mathcal{A}$ systems. Determining whether a given system is class $\mathcal{A}$ is numerically simple, as

one only needs to check the condition in (153). From a theoretical point of view, it would be relevant to offer a characterization of class $\mathcal{A}$ systems through properties of the system matrices. A detailed study to that end will be a digression from the main theme of the paper, and we intend to do it elsewhere. However, we show that scalar systems are included in class $\mathcal{A}$, thereby confirming that it is not empty.

Proposition 8.4 Let (A, Q, C, R) be a scalar system satisfying **(E.1)** (i.e., $Q, R > 0$ and $C \neq 0$.) Then, (A, Q, C, R) belongs to class $\mathcal{A}$.

Proof: Define the function $g : \mathbb{R}_+ \longmapsto \mathbb{R}_+$ by

$$g(X) = f_1(X) - X, \quad \forall X \geq 0 \tag{156}$$

where f_1 is the scalar Riccati operator. Under the assumptions, f_1 has only one fixed point in $\mathbb{R}_+$, P^*, the steady state solution. Thus, in the domain of interest,

$$g(X) = 0 \quad \textit{iff} \quad X = P^* \tag{157}$$

We note that $g(0) > 0$, and, by [22], there exists $\alpha_{P^*} \geq P^*$ sufficiently large, such that,

$$g(X) = f_1(X) - X < 0, \quad \forall X > \alpha_{P^*} \tag{158}$$

We now claim that

$$g(X) > 0, \quad 0 \leq X < P^* \quad \text{and} \quad g(X) < 0, \quad X > P^* \tag{159}$$

Indeed, if this was not true, then by (158) this would imply the existence of an interval in $\mathbb{R}_+$ not containing P^*, such that the sign of $g(\cdot)$ changes over this interval. This in turn would imply from the continuity of $g(\cdot)$, the existence of another solution to the equation $g(X) = 0$ on $\mathbb{R}_+$ other than P^*. Clearly, this contradicts with the hypothesis (see (157) and, hence, the claim in (159) holds.

Since $f_0(P^*) > P^*$, it then follows from (159) that, for $X \geq f_0(P^*)$,

$$f_1(X) - X = g(X) < 0 \tag{160}$$

thus showing that scalar systems belong to class $\mathcal{A}$. ■

As shown by Lemma 8.2, for class $\mathcal{A}$ systems the computation of the decay exponent of rare events can be greatly simplified. For general systems such a simplification may not be possible, however, the solution of the variational problem in Theorem 4.2 can still be made more efficient rather than searching haphazardly over the set of strings $\mathcal{S}^{P^*}$. The following proposition outlines a simple algorithm for solving the variational problems of interest leading to the decay exponents of rare events in general systems:

Proposition 8.5 Let $\Gamma \subset \mathbb{S}_+^N$ and define

$$k_\Gamma = \inf\{k \geq 0 \mid \mathcal{N}\left(f_0^k(P^*)\right) \in \Gamma\} \tag{161}$$

Define the set

$$\mathcal{J}_\Gamma = \{\mathcal{R} \in \mathcal{S}^{P^*} \mid \mathcal{N}(\mathcal{R}) \in \Gamma \ \text{ and } \ \text{len}(\mathcal{R}) \leq k_\Gamma\} \tag{162}$$

Then

$$\inf_{X \in \Gamma} I(X) = \inf_{\mathcal{R} \in \mathcal{J}_\Gamma} \pi(\mathcal{R}) \tag{163}$$

Proof: The proof is straight-forward and follows from the fact, that, $\inf_{X\in\Gamma} I(X) \leq k_\Gamma$ and so it suffices to look at strings of length k_Γ at most. ■

Remark 8.6 The conclusion of Proposition 8.5 is that, to find the decay exponent of a rare event Γ, one may compute k_Γ according to (161) and then the minimizing string can be found in the set $\mathcal{J}_\Gamma$ as constructed above.

9. A Scalar Example

We present a numerical study to demonstrate the efficiency of our approach over extensive Monte-Carlo type simulations to estimate the decay rate of rare events.

Consider a scalar system with parameters: $A = \sqrt{2}, C = Q = R = 1$. Solving the algebraic Riccati equation, $X = f_1(X)$, we obtain $P^* = 1 + \sqrt{2}$. By Proposition 8.4, we note that the system is of class $\mathcal{A}$. By Lemma 8.2, we then have for $M > 0$,

$$-\iota_+(M) \liminf_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}} \left(B_M^C(P^*)\right) \leq \limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}} \left(B_M^C(P^*)\right) \leq -\iota(M) \tag{164}$$

(note, we use the alternative MDP representation, (10).)

Now choose $M_1 = 40 - P^*$ and we estimate the decay rate of the rare event $B_{M_1}^C(P^*)$ as $\overline{\gamma} \to 1$. Using the definitions and that

$$P^* < f_0^3(P^*) < M_1 + P^* < f_0^4(P^*) \tag{165}$$

we note that

$$\iota(M_1) = \iota_+(M_1) = 4 \tag{166}$$

By (164), $B_{M_1}^C(P^*)$ is an I-continuity set (see text following (12)), hence the limit in (164) exists and we have

$$\lim_{\overline{\gamma}\uparrow 1} \frac{1}{\ln(1-\overline{\gamma})} \ln \mu^{\overline{\gamma}} \left(B_{M_1}^C(P^*)\right) = 4 \tag{167}$$

Our theory then predicts, that, for $\overline{\gamma}$ close to 1,

$$\mu^{\overline{\gamma}} \left(B_{M_1}^C(P^*)\right) \sim (1-\overline{\gamma})^4 \tag{168}$$

We now estimate the empirical decay rate of the event through extensive numerical simulations. We simulate different values of $\overline{\gamma}$, in the range $[.55, 1)$ with a step size of .05. For each such value of $\overline{\gamma}$, we obtained 10^4 samples from the invariant measure $\mu^{\overline{\gamma}}$ (this is needed as the event $\left(B_{M_1}^C(P^*)\right)$ becomes increasingly difficult to observe as $\overline{\gamma}$ approaches 1.) Obtaining a sample from $\mu^{\overline{\gamma}}$ is also numerically intensive, and we iterated the RRE 100 times to make sure the random covariance converged in distribution to $\mu^{\overline{\gamma}}$. Thus, a total of $9 \times 10^4 \times 10^2$ computations were involved. The resulting empirical cumulative distributions functions (cdf) are plotted in Fig. 1 (on the left). We note that, as $\overline{\gamma} \to 1$, the empirical invariant measures converge to δ_{P^*}, thus verifying Theorem 4.1.

To obtain the empirical decay rate of $B_{M_1}^C(P^*)$, for each $\overline{\gamma}$, we numerically estimate the quantity $\frac{\ln \mu^{\overline{\gamma}}\left(B_{M_1}^C(P^*)\right)}{\ln(1-\overline{\gamma})}$ from the empirical cdfs obtained above. This is plotted in Fig. 1 (on the right) as a function of $\overline{\gamma}$ (the solid line). The result agrees with our theoretical findings: (1) qualitatively, the rare event decays as a power law of $(1-\overline{\gamma})$; (2) the best decay exponent approaches 4 (which is theoretically established in (167)) as $\overline{\gamma}$ gets closer to 1; (3) even for $\overline{\gamma}$ much less than 1, the empirical decay rate is close to 4, justifying (168).

This example truly demonstrates the relevance of our theoretical findings. Even for the simple scalar case, a modest numerical estimation of the probabilities of rare events required computations of the order 10^7, whereas, our theoretical findings provide the best decay exponent by solving a much simpler variational problem.

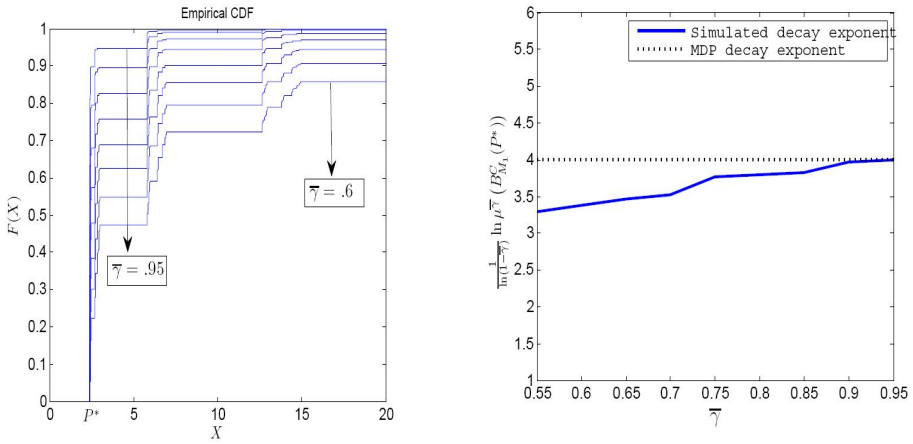


Fig. 1. Left: Weak convergence of (empirical) measures $\mu^{\overline{\gamma}}$ to δ_{P^*} as $\overline{\gamma} \to 1$. Right: Decay exponent of probability of the rare event $B^C_{M_1}(P^*)$.

10. Conclusions

The paper studies the RRE arising from the problem of Kalman filtering with intermittent observations. We show that for every $\overline{\gamma} > 0$ the conditional mean-squared error process is ergodic and the resulting family of invariant distributions $\{\mu^{\overline{\gamma}}\}$ (as they converge weakly to δ_{P^*} as $\overline{\gamma} \uparrow 1$) satisfies a MDP with good rate function I. The rate function I is completely characterized and the asymptotic decay rate of rare events characterized as solutions of deterministic variational problems. The intermediate results obtained are of independent interest and our methods are fairly general to be applicable to the analysis of more complex networked control systems (see, for example, [18]) and hybrid or switched systems.

11. Acknowledgements

Soummya Kar wishes to thank Kavita Ramanan of the Department of Mathematical Sciences, Carnegie Mellon University, for insightful discussions on large deviations of Markov processes.

Appendix A
Proofs of Proposition 3.6, Lemma 3.7

Proof of Proposition 3.6

Proof: Assertion (i) follows from the fact, that, $f_1(P^*) = P^*$, whereas Assertion (ii) is obvious from the iterated construction of the sequence $\{P_t\}_{t \in \mathbb{T}_+}$. We prove Assertion (iii) now.

Following Bucy ([22]), we define the set

$$S^- = \left\{ X \in \mathbb{S}_+^N \mid f_1(X) \preceq X \right\} \tag{169}$$

Under the assumptions of observability and controllability, it can be shown (see [22]), there exists $\beta \in \mathbb{R}_+$, such that,

$$\left\{ X \in \mathbb{S}_+^N \mid X \succeq \beta I \right\} \subset S^- \tag{170}$$

Now choose $\alpha_{P^0} \geq \beta$, such that, $P^0 \preceq \alpha_{P^0} I$. Then $\alpha_{P^0} \in S^-$ and it follows from the order-preserving property of the operators f_0 and f_1, that,

$$\mathcal{N}(\mathcal{R}) \preceq \mathcal{N}(\mathcal{R}_1) \tag{171}$$

where $\mathcal{R}_1 = (f_{i_1}, \cdots, f_{i_t}, \alpha_{P^0} I)$. Note that the claim is trivial for $t = 0$, so assume $t \geq 1$. Let $j_1 \leq \cdots \leq j_{\pi(\mathcal{R})}$ be the indices corresponding to f_0s in $\mathcal{R}$ from left to right. Consider the last segment of $\mathcal{R}_1$, i.e., the string $\left(f_0, f_1^{t-j_{\pi(\mathcal{R})}}, \alpha_{P^0} I\right)$.

We then have

$$f_0 \circ f_1^{t-j_{\pi(\mathcal{R})}} (\alpha_{P^0} I) \preceq f_0 (\alpha_{P^0} I) \tag{172}$$

This follows from the fact, that, $f_1 (\alpha_{P^0} I) \preceq \alpha_{P^0} I$, which implies the sequence $\{f_1^s (\alpha_{P^0} I)\}_{s \geq 0}$ is decreasing and we have $f_1^{t-j_{\pi(\mathcal{R})}} (\alpha_{P^0} I) \preceq \alpha_{P^0} I$. Since $\alpha_{P^0} \geq \beta$ and A is unstable, we have $f_0 (\alpha_{P^0} I) \succeq \alpha_{P^0} I \succeq \beta I$ and $f_0 (\alpha_{P^0} I) \subset S^-$. In particular, we have

$$\mathcal{N}(\mathcal{R}_1) \preceq f_{i_1} \circ \cdots f_{i_{j_{\pi(\mathcal{R})}-1}} (f_0 (\alpha_{P^0} I)) \tag{173}$$

using the order-preserving property of the functions f_0, f_1. In a similar way, we can repeat the above argument inductively starting with the string on the R.H.S. of (173) and arrive at $\mathcal{N}(\mathcal{R}) \preceq f_0^{\pi(\mathcal{R})} (\alpha_{P^0} I)$. The claim then follows from (171.) ■

Proof of Lemma 3.7

Proof: We use the following result on uniform convergence over compact sets of the Riccati iterates to P^*. Let $X_1, X_2 \in \mathbb{S}_+^N$. Then it can be shown (see Theorem 7.5 in [23])[8] that there exist constants $c_1, c_2 > 0$, such that,

$$\left\| f_1^t(X_1) - f_1^t(X_2) \right\| \leq c_1 e^{-c_2 t} \|X_1 - X_2\| \tag{174}$$

Taking $X_1 = X$, $X_2 = P^*$ and noting that $f_1(P^*) = P^*$ we have from the above

$$\left\| f_1^t(X) - P^* \right\| \leq c_1 e^{-c_2 t} \|X_1 - P^*\| \tag{175}$$

Thus for every compact subset $K \in \mathbb{S}_+^N$, there exists $\acute{t}_\varepsilon$, depending on K, such that,

$$\left\| f_1^t(X) - P^* \right\| \leq \varepsilon, \quad X \in K, \quad t \geq \acute{t}_\varepsilon \tag{176}$$

To prove the Lemma, we need to transform the above uniform convergence over compact sets to uniform convergence over the entire space $\mathbb{S}_+^N$. To this end, we note that for a observable and controllable system, the following uniform boundedness of Riccati iterates holds from arbitrary initial state $X \in \mathbb{S}_+^N$ (see, for example, Lemma 7.1 in [23]):

$$f^t(X) \leq \alpha_1 I, \quad t \geq N, \quad \forall X \in \mathbb{S}_+^N \tag{177}$$

where $\alpha_1 \in \mathbb{R}_+$ is a sufficiently large number.

Now consider the compact set

$$K_\alpha = \left\{ X \in \mathbb{S}_+^N \mid \|X\| \leq \alpha \right\} \tag{178}$$

From (177) we have

$$f^t(X) \in K_\alpha, \quad t \geq N, \quad \forall X \in \mathbb{S}_+^N \tag{179}$$

Now from (176) choose $\acute{t}_\varepsilon$, such that,

$$\left\| f_1^t(X) - P^* \right\| \leq \varepsilon, \quad X \in K_\alpha, \quad t \geq \acute{t}_\varepsilon \tag{180}$$

Then defining $t_\varepsilon = \acute{t}_\varepsilon + N$ we have from eqns. (179,180)

$$\left\| f_1^t(X) - P^* \right\| \leq \varepsilon, \quad t \geq t_\varepsilon \tag{181}$$

and the result follows. ■

APPENDIX B
STOCHASTIC BOUNDEDNESS OF ERROR COVARIANCES

Proof of Lemma 3.9

Proof: We consider the case of unstable A. For stable A, the proposition is trivial and follows from the fact, that, the unconditional variance of the state sequence reaches a steady state (hence bounded), and a suboptimal estimate $\widehat{x}_t \equiv 0$ for all t leads to pathwise boundedness of the corresponding error covariance. In fact, in this case even $\overline{\gamma} = 0$ leads to stochastic boundedness of the sequence $\{P_t\}$ from every initial condition.

The proof follows the same line of arguments used in Proposition 6 of [2], where the above claim was established for the case of invertible C. The key ingredients used there consisted of uniformly bounding the Riccati operator (in case C is invertible) and then estimating the probability that the random sequence $\{P_t\}_{t \in \mathbb{T}_+}$ exceeds a particular range by relating it to the length of the random time intervals between packet arrivals.

In the general case, as shown below, instead of bounding the one-step Riccati iterates, we bound N-step Riccati iterates for controllable and observable systems and then repeat the arguments in [2] with more generality.

[8]The result in [23] applies to time-variant system matrices under the assumptions of uniform complete observability and uniform complete controllability, which reduces to the observability and controllability for the time-invariant system matrices considered in this paper.

To this end, we start with the following result on boundedness of N-step Riccati iterates. It can be shown (see Lemma 7.1 in [23]) that for controllable and observable systems[9] the following holds:

$$f_1^t(X) \preceq \kappa I, \quad t \geq N, \quad X \in \mathbb{S}_N^+ \tag{182}$$

where $\alpha_2 \in \mathbb{R}_+$ is sufficiently large. In other words, the above states that an application of the Riccati operator more than N times in succession leads to a covariance bounded above by a specific constant, irrespective of the initial state.

Now, for $M \in \mathbb{T}_+$ and sufficiently large, define

$$k(M) = \min\left\{k \in \mathbb{T}_+ \mid \kappa_1 \alpha^{2k} + \|Q\| \frac{\alpha^{2k}-1}{\alpha^2-1} \geq M\right\} \tag{183}$$

where $\alpha = \|A\|$ and $\kappa_1 = \max\{\kappa, \|P_0\|\}$. Since A is unstable ($\alpha > 1$), it follows that $k(M) \to \infty$ as $N \to \infty$. To estimate the probability $\mathbb{P}^{\overline{\gamma},P_0}(\|P_t\| > M)$ for $t \in \mathbb{T}_+$, define the random time $\widetilde{t}$ by

$$\widetilde{t} = \max\{0 < s \leq t \mid \gamma_{s-r} = 1, \quad 1 \leq r \leq N\} \tag{184}$$

where the maximum of an empty set is taken to be zero. Thus, if $\widetilde{t} \neq 0$,[10] it denotes the time closest to t, such that, there were N successive packet arrivals in the time interval $[\widetilde{t}-N, \widetilde{t}-1]$. Then, using the above arguments, we have

$$\|P_{\widetilde{t}}\| \leq \kappa_1 \tag{185}$$

Indeed, if $\widetilde{t} = 0$, then $P_{\widetilde{t}} = P_0$ and (185) holds by the definition of κ_1. On the contrary, if $\widetilde{t} > 0$, we have by (182)

$$\|P_{\widetilde{t}}\| = \left\|f_1^N\left(P_{\widetilde{t}-N}\right)\right\| \leq \kappa \leq \kappa_1 \tag{186}$$

We then have

$$\|P_t\| = \left\|f_{\gamma_t-1} \circ \cdots \circ f_{\gamma_{\widetilde{t}}}(P_{\widetilde{t}})\right\| \leq \left\|f_0^{t-\widetilde{t}}(P_{\widetilde{t}})\right\| \leq \kappa_1 \alpha^{2(t-\widetilde{t})} + \|Q\| \sum_{k=1}^{t-\widetilde{t}} \alpha^{2k} = \kappa_1 \alpha^{2(t-\widetilde{t})} + \|Q\| \frac{\alpha^{2(t-\widetilde{t})}-1}{\alpha^2-1} \tag{187}$$

where we have used the fact, that, $f_1(X) \preceq f_0(X), \quad \forall\, X \in \mathbb{S}_+^N$. It then follows from the above and (183), that,

$$\mathbb{P}^{\overline{\gamma},P_0}(\|P_t\| > M) \leq \mathbb{P}_{\overline{\gamma},P_0}\left(t - \widetilde{t} \geq k(M)\right) \tag{188}$$

We now estimate the probability $\mathbb{P}_{\overline{\gamma},P_0}\left(t-\widetilde{t}=k\right)$.

First, consider the case $\widetilde{t} \neq 0$. On the event $\widetilde{t} \neq 0$, it is not hard to see that the following events are equal:

$$\left\{t-\widetilde{t}=k\right\} = \{\gamma_{t-k-r} = 1, \quad 1 \leq r \leq N\} \bigcap_{s=t-k+1}^{t} \{\gamma_{s-r} = 1, \quad 1 \leq r \leq N\}^c \tag{189}$$

It then follows by elementary manipulations and the independence of packet arrivals

$$\begin{aligned}
\mathbb{P}_{\overline{\gamma},P_0}\left(t-\widetilde{t}=k\right) &= \mathbb{P}_{\overline{\gamma},P_0}\left(\{\gamma_{t-k-r} = 1, \quad 1 \leq r \leq N\} \bigcap_{s=t-k+1}^{t} \{\gamma_{s-r} = 1, \quad 1 \leq r \leq N\}^c\right) \\
&\leq \mathbb{P}_{\overline{\gamma},P_0}\left(\{\gamma_{t-k-r} = 1, \quad 1 \leq r \leq N\} \bigcap_{i=1}^{\lfloor \frac{k}{N} \rfloor} \left\{\gamma_{t-k+(i-1)N-1+r} = 1, \quad 1 \leq r \leq N\right\}^c\right) \\
&= \mathbb{P}_{\overline{\gamma},P_0}\left(\{\gamma_{t-k-r} = 1, \quad 1 \leq r \leq N\}\right) \prod_{i=1}^{\lfloor \frac{k}{N} \rfloor} \mathbb{P}_{\overline{\gamma},P_0}\left(\left\{\gamma_{t-k+(i-1)N-1+r} = 1, \quad 1 \leq r \leq N\right\}^c\right) \\
&= \overline{\gamma}^N \prod_{i=1}^{\lfloor \frac{k}{N} \rfloor} \left(1-\overline{\gamma}^N\right) \\
&\leq \left(1-\overline{\gamma}^N\right)^{\lfloor \frac{k}{N} \rfloor}
\end{aligned} \tag{190}$$

[9]The result in [23] applies to time-variant system matrices under the assumptions of uniform complete observability and uniform complete controllability, which reduces to the observability and controllability for the time-invariant system matrices considered in this paper.

[10]Note that, if not zero, $\widetilde{t} \geq N$.

On the event $\widetilde{t}=0$, using a similar set of arguments, we can show

$$\mathbb{P}_{\overline{\gamma},P_0}\left(t-\widetilde{t}=k\right)\leq\left(1-\overline{\gamma}^N\right)^{\lfloor\frac{k}{N}\rfloor} \tag{191}$$

We thus have the upper bound (possibly loose, but sufficient for our purpose)

$$\mathbb{P}_{\overline{\gamma},P_0}\left(t-\widetilde{t}\geq k(M)\right)=\sum_{k=k(M)}^{\infty}\mathbb{P}_{\overline{\gamma},P_0}\left(t-\widetilde{t}=k\right)\leq\sum_{k=k(M)}^{\infty}\left(1-\overline{\gamma}^N\right)^{\lfloor\frac{k}{N}\rfloor}\leq\sum_{k=k(M)}^{\infty}\left(1-\overline{\gamma}^N\right)^{\frac{k}{N}-1} \tag{192}$$

Rearranging and summing the geometric series above we have

$$\mathbb{P}_{\overline{\gamma},P_0}\left(t-\widetilde{t}\geq k(M)\right)\leq\frac{1}{1-\overline{\gamma}^N}\sum_{k=k(M)}^{\infty}\left[\left(1-\overline{\gamma}^N\right)^{1/N}\right]^k=\frac{1}{1-\overline{\gamma}^N}\frac{\left[\left(1-\overline{\gamma}^N\right)^{1/N}\right]^{k(M)}}{1-\left(1-\overline{\gamma}^N\right)^{1/N}} \tag{193}$$

From eqns. (188,193) we have for all t and sufficiently large M

$$\mathbb{P}^{\overline{\gamma},P_0}\left(\|P_t\|>M\right)\leq\frac{1}{1-\overline{\gamma}^N}\frac{\left[\left(1-\overline{\gamma}^N\right)^{1/N}\right]^{k(M)}}{1-\left(1-\overline{\gamma}^N\right)^{1/N}} \tag{194}$$

Since $\overline{\gamma}>0$ and $k(M)\to\infty$ as $M\to\infty$, it follows from the above

$$\lim_{M\to\infty}\sup_{t\in\mathbb{T}_+}\mathbb{P}^{\overline{\gamma},P_0}\left(\|P_t\|>M\right)=0 \tag{195}$$

Thus $\{P_t\}_{t\in\mathbb{T}_+}$ is s.b. (for all initial conditions P_0) for every $\overline{\gamma}>0$ and hence, the Lemma follows.

■

APPENDIX C
PROOF OF THEOREM 5.2

Proof: For $\mathbf{X}=\{X_1,\cdots,X_n\}\in\bigotimes_{i=1}^n\mathbb{S}_+^N$, we have

$$\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}=\mathbf{X}\right)=\mathbb{P}^{\overline{\gamma},P_0}\left(P_{t_1}=X_1\right)\prod_{i=1}^{n-1}\mathbb{P}^{\overline{\gamma},P_0}\left(P_{t_{i+1}}=X_{i+1}\mid P_{t_i}=X_i\right) \tag{196}$$

which follows from the Markov property. Clearly, from the above we have

$$\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}=\mathbf{X}\right)=0,\quad\mathbf{X}\notin\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right) \tag{197}$$

Also, if $\mathbf{X}\in\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)$, it follows from (196) through simple manipulations and the independence of the packet dropouts

$$\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}=\mathbf{X}\right)=\sum_{\overline{\mathcal{R}}\in\mathcal{S}^{P_0}_{t_1,\cdots,t_n}(\mathbf{X})}(1-\overline{\gamma})^{\pi(\overline{\mathcal{R}})}\overline{\gamma}^{t_n-\pi(\overline{\mathcal{R}})} \tag{198}$$

By manipulating each term on the R.H.S. above, we have

$$\begin{aligned}\lim_{\overline{\gamma}\uparrow1}\frac{1}{\ln(1-\overline{\gamma})}\ln\left((1-\overline{\gamma})^{\pi(\overline{\mathcal{R}})}\overline{\gamma}^{t_n-\pi(\overline{\mathcal{R}})}\right)&=\lim_{\overline{\gamma}\uparrow1}\frac{1}{\ln(1-\overline{\gamma})}\left[\pi(\overline{\mathcal{R}})\ln(1-\overline{\gamma})+\left(t_n-\pi(\overline{\mathcal{R}})\right)\ln\overline{\gamma}\right]\\&=\pi(\overline{\mathcal{R}})+\left(t_n-\pi(\overline{\mathcal{R}})\right)\lim_{\overline{\gamma}\uparrow1}\frac{\ln\overline{\gamma}}{\ln(1-\overline{\gamma})}\\&=\pi(\overline{\mathcal{R}})\end{aligned} \tag{199}$$

It then follows from Proposition 3.10,eqns. (198,199) that

$$\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}=\mathbf{X}\right)=\min_{\overline{\mathcal{R}}\in\mathcal{S}^{P_0}_{t_1,\cdots,t_n}(\mathbf{X})}\pi(\overline{\mathcal{R}})=\ell(\mathbf{X}) \tag{200}$$

(the case $\mathbf{X}\notin\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)$ is absorbed above by using the convention, that, the minimum of an empty set is ∞.)

Now consider $B\in\mathcal{B}\left(\bigotimes_{i=1}^n\mathbb{S}_+^N\right)$. If $B\cap\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)=\phi$, then the claim is obvious. Hence, assume $B\cap\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)\neq\phi$ (this intersection is necessarily finite) and note that

$$\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}\in B\right)=\sum_{\mathbf{X}\in B\cap\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)}\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}=\mathbf{X}\right) \tag{201}$$

It then follows by Proposition 3.10

$$\begin{aligned}\lim_{\overline{\gamma}\uparrow 1}\frac{1}{\ln(1-\overline{\gamma})}\ln\left(\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}\in B\right)\right) &= \min_{\mathbf{X}\in B\cap\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)}\lim_{\overline{\gamma}\uparrow 1}\frac{1}{\ln(1-\overline{\gamma})}\ln\left(\mathbb{P}^{\overline{\gamma},P_0}\left((P_{t_1},\cdots,P_{t_n})^{\overline{\gamma}}=\mathbf{X}\right)\right)\\ &= \min_{\mathbf{X}\in B\cap\mathcal{N}\left(\mathcal{S}^{P_0}_{t_1,\cdots,t_n}\right)}\ell(\mathbf{X})\\ &= \inf_{\mathbf{X}\in B} I^{t_1,\cdots,t_n}(\mathbf{X}) \qquad (202)\end{aligned}$$

which establishes the Theorem. ∎

APPENDIX D
PROOFS OF RESULTS IN SECTION 6

Proof of Proposition 6.1

Proof: That $I_L(\cdot)$ is lower semicontinuous follows from its definition (see, for example, [24].) Now for $a\in\mathbb{R}_+$ consider the level set $K_a=\{X\in\mathbb{S}^N_+ \mid I_L(X)\leq a\}$. By lower semicontinuity we know that K_a is closed. We now show that K_a is bounded and hence compact. To this end, we note that for all $b\in\mathbb{R}_+$

$$\left\{Y\in\mathbb{S}^N_+ \mid I(Y)\leq b\right\}\subset\left\{Y\in\mathbb{S}^N_+ \mid Y\preceq f_0^{\lceil b\rceil}(\alpha_{P^*}I)\right\} \qquad (203)$$

for some constant $\alpha_{P^*}\in\mathbb{R}_+$, which can be chosen independent of b. Indeed, $I(Y)\leq b$ implies that $\mathcal{S}^{P^*}(Y)$ is non-empty and

$$I(Y)=\inf_{\mathcal{R}\in\mathcal{S}^{P^*}(Y)}\pi(\mathcal{R})\leq b \qquad (204)$$

Since $\pi(\cdot)$ takes on integral values only, the infimum above is attained and there exists $\mathcal{R}\in\mathcal{S}^{P^*}(Y)$ with $\pi(\mathcal{R})\leq b$. Then, from Proposition 3.6, there exists $\alpha_{P^*}\in\mathbb{R}_+$ (depending on P^* only), such that

$$Y=\mathcal{N}(\mathcal{R})\preceq f_0^{\pi(\mathcal{R})}(\alpha_{P^*}I)\preceq f_0^{\lceil b\rceil}(\alpha_{P^*}I) \qquad (205)$$

This verifies the claim in (203).

Now consider $\varepsilon_1>0$. For $X\in\mathbb{S}^N_+$, the sequence $\inf_{Y\in B_\varepsilon(X)}I(Y)$ is non-decreasing w.r.t. ε and hence $X\in K_a$ implies

$$\inf_{Y\in B_{\varepsilon_1}(X)}I(Y)\leq a \qquad (206)$$

Since $I(\cdot)$ takes on integral values the infimum is attained and there exists $Y(X)\in\mathbb{S}^N_+$, such that, $I(Y(X))\leq a$. From (203) we then have

$$Y(X)\preceq f_0^{\lceil a\rceil}(\alpha_{P^*}I)\implies\|Y(X)\|\leq\left\|f_0^{\lceil a\rceil}(\alpha_{P^*}I)\right\| \qquad (207)$$

Since $Y(X)\in B_{\varepsilon_1}(X)$ we have

$$\|X\|\leq\|Y(X)\|+\varepsilon_1\leq\left\|f_0^{\lceil a\rceil}(\alpha_{P^*}I)\right\|+\varepsilon_1 \qquad (208)$$

We thus note that

$$K_a\subset\left\{Z\in\mathbb{S}^N_+ \mid \|Z\|\leq\left\|f_0^{\lceil a\rceil}(\alpha_{P^*}I)\right\|+\varepsilon_1\right\} \qquad (209)$$

which verifies the boundedness of K_a. Hence the level sets K_a are closed and bounded for all $a\in\mathbb{R}_+$, establishing the goodness of $I_L(\cdot)$.

For Assertion (ii), we note that for arbitrary $\varepsilon>0$,

$$\inf_{Y\in\overline{B_\varepsilon(X)}}I(Y)\leq\inf_{Y\in B_\varepsilon(X)}I(Y)\leq\inf_{Y\in\overline{B_{\varepsilon/2}(X)}}I(X) \qquad (210)$$

The assertion then follows by passing to the limit as $\varepsilon\to 0$ on each side.

For Assertion (iii), note that, in general we have for arbitrary $\varepsilon>0$, $I(X)\geq\inf_{Y\in B_\varepsilon(X)}I(X)$ and by passing to the limit it follows

$$I(X)\geq\lim_{\varepsilon\to 0}\inf_{Y\in B_\varepsilon(X)}I(X)=I_L(X) \qquad (211)$$

This immediately gives for any set $\Gamma\in\mathcal{B}(\mathbb{S}^N_+)$

$$\inf_{X\in\Gamma}I(X)\geq\inf_{X\in\Gamma}I_L(X) \qquad (212)$$

For the reverse inequality when Γ is open, consider $X\in\Gamma$. Then there exists $\varepsilon_1>0$ (depending on X) such that, for every $0<\varepsilon<\varepsilon_1$, the open ball $B_\varepsilon(X)\in\Gamma$. It then follows

$$\inf_{Y\in B_\varepsilon(X)}I(Y)\geq\inf_{Y\in\Gamma}I(Y),\quad 0<\varepsilon<\varepsilon_1 \qquad (213)$$

Taking the limit on both sides we have

$$\inf_{Y\in\Gamma} I(Y) \leq \lim_{\varepsilon\to 0} \inf_{Y\in B_\varepsilon(X)} I(Y) = I_L(X) \tag{214}$$

Thus for every $X \in \Gamma$, we have

$$I_L(X) \geq \inf_{Y\in\Gamma} I(Y) \tag{215}$$

Taking the infimum over $X \in \Gamma$ on the L.H.S. gives the required inequality

$$\inf_{X\in\Gamma} I_L(X) \geq \inf_{X\in\Gamma} I(X) \tag{216}$$

and the result follows.

We now prove Assertion (iv). By the notion of limits involving continuous arguments, it suffices to show that for every sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ with

$$\lim_{n\to\infty} \varepsilon_n = 0, \quad 0 < \varepsilon_n \leq 1 \ \forall n \tag{217}$$

we have

$$\lim_{n\to\infty} \inf_{Y\in\overline{K_{\varepsilon_n}}} I_L(Y) = \inf_{Y\in K} I_L(Y) \tag{218}$$

To this end consider such a sequence $\{\varepsilon_n\}$ and assume on the contrary, that (218) is not satisfied. Since $K \subset \overline{K_{\varepsilon_n}}$, we clearly have

$$\lim_{n\to\infty} \inf_{Y\in\overline{K_{\varepsilon_n}}} I_L(Y) \leq \inf_{Y\in K} I_L(Y) \tag{219}$$

Thus the hypothesis that (218) is not satisfied implies

$$\lim_{n\to\infty} \inf_{Y\in\overline{K_{\varepsilon_n}}} I_L(Y) < \inf_{Y\in K} I_L(Y) \tag{220}$$

We note that the sets $\overline{K_{\varepsilon_n}}$ are compact for all n and since a lower semicontinuous function attains its minimum over a compact set, for every $n \in \mathbb{N}$, there exists $X_n \in \overline{K_{\varepsilon_n}}$, such that

$$\inf_{Y\in\overline{K_{\varepsilon_n}}} I_L(Y) = I_L(X_n) \tag{221}$$

Similarly, there exists $Y^* \in K$, such that,

$$\inf_{Y\in K} I_L(Y) = I_L(Y^*) \tag{222}$$

We note that $\{I_L(X_n)\}$ is a non decreasing sequence (hence the limit exists) and

$$\lim_{n\to\infty} I_L(X_n) < I_L(Y^*) \tag{223}$$

Also, given $Z \in \mathbb{S}_+^N$, it follows from the continuity of the metric, that the function $d^Z : \mathbb{S}_+^N \longmapsto \mathbb{R}_+$ given by

$$d^Z(Y) = \|Y - Z\|, \quad \forall Y \in \mathbb{S}_+^N \tag{224}$$

is continuous and hence attains its minimum over a compact set. Thus for every $n \in \mathbb{N}$, there exists $Y_n \in K$, such that

$$\inf_{Y\in K} \|X_n - Y\| = \|X_n - Y_n\| \leq \varepsilon_n \tag{225}$$

where the last inequality follows from the fact that $X_n \in \overline{K_{\varepsilon_n}}$.

We note that the sequence $\{Y_n\}$ belongs to the compact set K and hence there exists a subsequence $\{Y_{n_k}\}_{k\in\mathbb{N}}$ in K, which converges to some $\acute{Y} \in K$, i.e.,

$$\lim_{k\to\infty} Y_{n_k} = \acute{Y} \tag{226}$$

Now consider the sequence $\{X_{n_k}\}_{k\in\mathbb{N}}$. We then have

$$\left\|X_{n_k} - \acute{Y}\right\| \leq \|X_{n_k} - Y_{n_k}\| + \left\|Y_{n_k} - \acute{Y}\right\| \leq \varepsilon_{n_k} + \left\|Y_{n_k} - \acute{Y}\right\| \tag{227}$$

Taking the limit as $k \to \infty$ we obtain

$$\lim_{k\to\infty} X_{n_k} = \acute{Y} \tag{228}$$

We then have from the lower semicontinuity of $I_L(\cdot)$ and (223)

$$I_L(\acute{Y}) \leq \liminf_{k\to\infty} I_L(X_{n_k}) < I_L(Y^*) \tag{229}$$

This contradicts the fact that Y^* is the minimizer of $I_L(\cdot)$ over K and we conclude that (218) holds. Hence the result follows. ■

Proof of Lemma 6.5

Proof: The case $\ell(\mathcal{F}) = 0$ is trivial as the assertion follows by choosing an arbitrary positive $t_{\mathcal{F}}$.

We consider the case $\ell(\mathcal{F}) \geq 1$. We use an inductive argument and it suffices to show that for every $1 \leq i \leq \ell(\mathcal{F})$ there exists positive $t^i_{\mathcal{F}} \in \mathbb{T}_+$, such that, for $\mathcal{R} \in \mathcal{U}(\mathcal{F})$ with $\text{len}(\mathcal{R}) \geq t^i_{\mathcal{F}}$, we have

$$\pi\left(\mathcal{R}^{t^i_{\mathcal{F}}}\right) \geq i \tag{230}$$

We start with $i = 1$. Assume on the contrary, that there is no such $t^1_{\mathcal{F}} \in \mathbb{T}_+$ with the above postulated properties. Since $\mathcal{U}(\mathcal{F})$ is non-empty, by Proposition 3.6 Assertion (i), there exists $t_0 \in \mathbb{T}_+$, such that

$$\mathcal{S}^{P^*}_t \cap \mathcal{U}(\mathcal{F}) \neq \phi \tag{231}$$

Thus the non-existence of $t^1_{\mathcal{F}}$, implies, that for every $t \geq t_0$, there exists a string $\mathcal{R}_t \in \mathcal{U}(\mathcal{F})$, with $\text{len}(\mathcal{R}_t) \geq t$, such that, $\pi\left(\mathcal{R}^t_t\right) = 0$. Such a string $\mathcal{R}_t$ is then necessarily of the form:

$$\mathcal{R} = \left(f^t_1, f_{i_1}, f_{i_2}, \cdots, f_{i_{\text{len}(\mathcal{R}_t)-t}}, P^*\right) \tag{232}$$

where $i_1, \cdots, i_{\text{len}(\mathcal{R}_t)-t} \in \{0, 1\}$. Thus denoting

$$X_t = f_{i_1} \circ f_{i_2} \circ \cdots \circ f_{i_{\text{len}(\mathcal{R}_t)-t}}(P^*) \tag{233}$$

we note that

$$\mathcal{N}(\mathcal{R}_t) = f^t_1(X_t) \tag{234}$$

Now consider the sequence $\{\mathcal{R}_t\}_{t \geq t_0}$ of such strings as $t \to \infty$. Let $\varepsilon > 0$ be an arbitrary positive number. By Proposition 3.7, the uniform convergence of Riccati iterates implies there exists $t_\varepsilon \geq N$, such that, for every $X \in \mathbb{S}^N_+$,

$$\left\|f^t_1(X) - P^*\right\| \leq \varepsilon, \quad t \geq t_\varepsilon \tag{235}$$

(we emphasize that the constant t_ε can be chosen independently of X.) Then defining $\acute{t}_\varepsilon = \max(t_0, t_\varepsilon)$, it follows from (234) that for $t \geq \acute{t}_\varepsilon$

$$\|\mathcal{N}(\mathcal{R}_t) - P^*\| = \left\|f^t_1(X_t) - P^*\right\| \leq \varepsilon \tag{236}$$

Since $\varepsilon > 0$ above was arbitrary, it follows that the sequence $\{\mathcal{N}(\mathcal{R}_t)\}_{t \geq t_0}$ of numerical values converges to P^* as $t \to \infty$. However, by construction, the sequence $\{\mathcal{N}(\mathcal{R}_t)\}_{t \geq t_0}$ belongs to the set $\mathcal{F}$ and we conclude that P^* is a limit point of the set $\mathcal{F}$. Since $\mathcal{F}$ is closed, we have $P^* \in \mathcal{F}$. It then follows

$$\left\{\mathcal{R} \in \mathcal{S}^{P^*} \mid \mathcal{N}(\mathcal{R}) = P^*\right\} \subset \mathcal{U}(\mathcal{F}) \tag{237}$$

Hence, in particular, the string $f_1(P^*) \in \mathcal{U}(\mathcal{F})$. The fact, that $\pi(f_1(P^*)) = 0$ then contradicts the hypothesis $\ell(\mathcal{F}) \geq 1$.

Thus by contradiction we establish that if $\ell(\mathcal{F}) \geq 1$, there exists $t^1_{\mathcal{F}}$ satisfying the properties in (230) for $i = 1$. Note that, if $\ell(\mathcal{F}) = 1$, this step completes the proof of the Lemma. In the general case, i.e., to establish (230) for all $1 \leq i \leq \ell(\mathcal{F})$ we need the following additional steps.

We can now assume $\ell(\mathcal{F}) \geq 2$. We assume on the contrary that the claim in (230) does not hold for all $1 \leq i \leq \ell(\mathcal{F})$. By the previous arguments, the claim clearly holds for $i = 1$. Then, let $1 \leq k < \ell(\mathcal{F})$ be the largest integer such that the claim in (230) holds for all $1 \leq i \leq k$. The hypothesis $k < \ell(\mathcal{F})$ implies there exists no $t^{k+1}_{\mathcal{F}} \in \mathbb{T}_+$ satisfying the claim in (230) for $i = k + 1$. Since the claim holds for $i = k$, there exists $t^k_{\mathcal{F}} \in \mathbb{T}_+$, such that for all $\mathcal{R} \in \mathcal{U}(\mathcal{F})$ with $\text{len}(\mathcal{R}) \geq t^k_{\mathcal{F}}$, we have $\pi\left(\mathcal{R}^{t^k_{\mathcal{F}}}\right) \geq k$. The non-existence of $t^{k+1}_{\mathcal{F}}$ and (231) implies, that for every $t \geq t_0$, there exists a string $\mathcal{R}_t \in \mathcal{U}(\mathcal{F})$, with $\text{len}(\mathcal{R}_t) \geq t$, such that,

$$\pi\left(\mathcal{R}^t_t\right) < k + 1 \tag{238}$$

We now study the structure of the strings $\mathcal{R}_t$ for sufficiently large t. To this end, define $\acute{t}_0 = \max(t_0, t^k_{\mathcal{F}})$. Then by the existence of $t^k_{\mathcal{F}}$ and (238) it follows that $\pi\left(\mathcal{R}^t_t\right) = k$ for $t \geq \acute{t}_0$. Hence for $t \geq \acute{t}_0$, $\mathcal{R}_t$ is necessarily of the form:

$$\mathcal{R}_t = \left(f_{i_1}, \cdots, f_{i_{t^k_{\mathcal{F}}}}, f_1^{t-t^k_{\mathcal{F}}}, f_{j_1}, f_{j_2}, \cdots, f_{j_{\text{len}(\mathcal{R}_t)-t}}, P^*\right) \tag{239}$$

where $i_1, \cdots, i_{t^k_{\mathcal{F}}} \in \{0, 1\}$, such that, $\pi\left(\mathcal{R}^{t^k_{\mathcal{F}}}\right) = k$ and $j_1, \cdots, j_{\text{len}(\mathcal{R}_t)-t} \in \{0, 1\}$.

Now consider the sequence $\{\mathcal{R}_t\}_{t \geq \acute{t}_0}$ and define the set $\mathcal{J}$ by $\mathcal{J} = \{\mathcal{R}_t, \ \ t \geq \acute{t}_0\}$. Also, define $\mathcal{J}_1 = \{\mathcal{R} \in \mathcal{S}^{P^*}_{t^k_{\mathcal{F}}} \mid \pi(\mathcal{R}) = k\}$. Consider the function $\Lambda^{t^k_{\mathcal{F}}} : \mathcal{J} \longmapsto \mathcal{J}_1$ by

$$\Lambda^{t^k_{\mathcal{F}}}(\mathcal{R}) = \mathcal{R}^{t^k_{\mathcal{F}}}, \quad \forall\ \mathcal{R} \in \mathcal{J} \tag{240}$$

Since the cardinality of $\mathcal{J}_1$ is finite and $\mathcal{J}$ is countably infinite, there exists $\mathcal{R}^* \in \mathcal{J}_1$, such that, the set $\left(\Lambda^{t^k_{\mathcal{F}}}\right)^{-1}(\{\mathcal{R}^*\})$ is

countably infinite. This in turn implies, that we can extract a subsequence $\{\mathcal{R}_{t_m}\}_{m\geq 0}$ from the sequence $\{\mathcal{R}_t\}_{t\geq \acute{t}_0}$, such that,

$$\mathcal{R}_{t_m}^{t_{\mathcal{F}}^k} = \mathcal{R}^*, \quad \forall m \geq 0 \tag{241}$$

In other words, if $\mathcal{R}^*$ is represented by $\mathcal{R}^* = \left(f_{i_1^*}, \cdots, f_{i^*_{t_{\mathcal{F}}^k}}, P^*\right)$ for some fixed $i_1^*, \cdots, i^*_{t_{\mathcal{F}}^k} \in \{0,1\}$, for every m, the string $\mathcal{R}_{t_m}$ is of the form:

$$\mathcal{R}_{t_m} = \left(f_{i_1^*}, \cdots, f_{i^*_{t_{\mathcal{F}}^k}}, f_1^{t_m - t_{\mathcal{F}}^k}, f_{j_1}, f_{j_2}, \cdots, f_{j_{\text{len}(\mathcal{R}_{t_m})-t}}, P^*\right) \tag{242}$$

where $j_1, \cdots, j_{\text{len}(\mathcal{R}_{t_m})-t} \in \{0,1\}$ are arbitrary. Denoting

$$X_m = f_{j_1} \circ f_{j_2} \circ \cdots \circ f_{j_{\text{len}(\mathcal{R}_{t_m})-t}}(P^*), \quad \forall m \tag{243}$$

we have

$$\mathcal{N}(\mathcal{R}_{t_m}) = f_{i_1^*} \circ \cdots \circ f_{i^*_{t_{\mathcal{F}}^k}} \left(f_1^{t_m - t_{\mathcal{F}}^k}(X_m)\right) \tag{244}$$

Since $t_m \to \infty$ as $m \to \infty$, using Proposition 3.7 in a similar way, we have

$$\lim_{m\to\infty} f_1^{t_m - t_{\mathcal{F}}^k}(X_m) = P^* \tag{245}$$

Noting that the function $f_{i_1^*} \circ \cdots \circ f_{i^*_{t_{\mathcal{F}}^k}} : \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ is continuous (being the composition of continuous functions), we then have

$$\begin{aligned}
\lim_{m\to\infty} \mathcal{N}(\mathcal{R}_{t_m}) &= \lim_{m\to\infty} f_{i_1^*} \circ \cdots \circ f_{i^*_{t_{\mathcal{F}}^k}} \left(f_1^{t_m - t_{\mathcal{F}}^k}(X_m)\right) \\
&= f_{i_1^*} \circ \cdots \circ f_{i^*_{t_{\mathcal{F}}^k}} \left(\lim_{m\to\infty} f_1^{t_m - t_{\mathcal{F}}^k}(X_m)\right) \\
&= f_{i_1^*} \circ \cdots \circ f_{i^*_{t_{\mathcal{F}}^k}} (P^*) \\
&= \mathcal{N}(\mathcal{R}^*)
\end{aligned} \tag{246}$$

Thus the sequence $\{\mathcal{N}(\mathcal{R}_{t_m})\}_{m\geq 0}$ in $\mathcal{F}$ converges to $\mathcal{N}(\mathcal{R}^*)$ as $m \to \infty$. Hence, $\mathcal{N}(\mathcal{R}^*) \in \mathcal{F}$ as $\mathcal{F}$ is closed and $\mathcal{N}(\mathcal{R}^*)$ is a limit point of $\mathcal{F}$. This in turn implies $\mathcal{R}^* \in \mathcal{U}(\mathcal{F})$. Since $\pi(\mathcal{R}^*) = k$, this contradicts the hypothesis $k < \ell(\mathcal{F})$ and the claim in (230) holds for all $1 \leq i \leq \ell(\mathcal{F})$. This establishes the Lemma.

■

Proof of Lemma 6.8

Proof: Let $a > 0$ be arbitrary and $z \in \mathbb{N}$ be such that $z \geq a$. From Proposition 3.6 Assertion (iii), there exists $\alpha_{P^*} \in \mathbb{R}_+$, such that,

$$\mathcal{N}(\mathcal{R}) \preceq f_0^{\pi(\mathcal{R})}(P^*), \quad \forall \mathcal{R} \in \mathcal{S}^{P^*} \tag{247}$$

Define $b \in \mathbb{R}_+$ such that $\|f_0^z(P^*)\| < b$ and consider the compact set $K_a = \left\{X \in \mathbb{S}_+^N \mid \|X\| \leq b\right\}$. Also define the closed set, $\mathcal{F}_b$, by $\mathcal{F}_b = \left\{X \in \mathbb{S}_+^N \mid \|X\| \geq b\right\}$. As per Lemma 6.5, define the set $\mathcal{U}(\mathcal{F}_b)$ as

$$\mathcal{U}(\mathcal{F}_b) = \left\{\mathcal{R} \in \mathcal{S}^{P^*} \mid \mathcal{N}(\mathcal{R}) \in \mathcal{F}_b\right\} \tag{248}$$

We then have the following inclusion:

$$\mathcal{U}(\mathcal{F}_b) \subset \left\{\mathcal{R} \in \mathcal{S}^{P^*} \mid \pi(\mathcal{R}) \geq z\right\} \tag{249}$$

and hence

$$\ell(\mathcal{F}_b) = \inf_{\mathcal{R}\in\mathcal{U}(\mathcal{F}_b)} \pi(\mathcal{R}) \geq z \tag{250}$$

Since $\mathcal{F}_b$ is closed, by Lemma 6.5 there exists $t_{\mathcal{F}_b} \in \mathbb{T}_+$, such that,

$$\pi(\mathcal{R}^{t_{\mathcal{F}_b}}) \geq z, \quad \forall \mathcal{R} \in \mathcal{U}(\mathcal{F}_b) \tag{251}$$

To estimate the probability $\mu^{\overline{\gamma}}(K_a^C)$, we now follow a similar set of arguments as used in Lemma 6.7. First note that we have by weak convergence:

$$\mu^{\overline{\gamma}}(K_a^C) \leq \liminf_{t\to\infty} \mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in K_a^C\right) \leq \liminf_{t\to\infty} \mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in \mathcal{F}_b\right) \tag{252}$$

For $t \in \mathbb{T}_+$ define the sets $\mathcal{J}_t^{P^*} = \mathcal{S}_t^{P^*} \cap \mathcal{U}(\mathcal{F}_b)$. For $t \geq t_{\mathcal{F}_b}$ we have (see also (105))

$$\mathbb{P}^{\overline{\gamma},P^*}\left(P_t^{\overline{\gamma},P^*} \in \mathcal{F}_b\right) = \sum_{\mathcal{R}\in\mathcal{J}_t^{P^*}} (1-\overline{\gamma})^{\pi(\mathcal{R})}\overline{\gamma}^{t-\pi(\mathcal{R})} \leq \binom{t_{\mathcal{F}_b}}{\ell(\mathcal{F}_b)}(1-\overline{\gamma})^{\ell(\mathcal{F}_b)} \leq \binom{t_{\mathcal{F}_b}}{\ell(\mathcal{F}_b)}(1-\overline{\gamma})^{z} \tag{253}$$

A familiar set of arguments as in Lemma 6.7 yields the following from eqns. (252,253):

$$\mu^{\overline{\gamma}}(K_a^C) \leq \binom{t_{\mathcal{F}_b}}{\ell(\mathcal{F}_b)}(1-\overline{\gamma})^{z}, \quad \forall\overline{\gamma} \tag{254}$$

from which we obtain

$$\limsup_{\overline{\gamma}\uparrow 1} -\frac{1}{\ln(1-\overline{\gamma})}\mu^{\overline{\gamma}}(K_a^C) \leq -z \leq -a \tag{255}$$

Thus for every $a > 0$ there exists a compact set K_a such that (111) is satisfied and the Lemma follows. ■

APPENDIX E
PROOF OF LEMMA 8.2

Proof: By Theorem 4.2 it suffices to show that for all $M > 0$,

$$\inf_{X\in B_M^C(P^*)} I(X) = \iota(M) \tag{256}$$

$$\inf_{X\in \overline{B}_M^C(P^*)} I(X) = \iota_+(M) \tag{257}$$

We prove (256) only, the proof of (257) being similar.

The class $\mathcal{A}$ assumption implies (through the same line of arguments as in Proposition 3.6 Assertion (iii)) that

$$\mathcal{N}(\mathcal{R}) \preceq f_0^{\pi(\mathcal{R})}(P^*), \quad \forall\ \mathcal{R} \in \mathcal{S}^{P^*} \tag{258}$$

By definition it follows that $f_0^{\iota(M)}(P^*) \in B_M^C(P^*)$ and hence

$$\inf_{X\in B_M^C(P^*)} I(X) \leq \iota(M) \tag{259}$$

Now assume on the contrary that (256) does not hold. Then by (259) we have

$$\inf_{X\in B_M^C(P^*)} I(X) < \iota(M) \tag{260}$$

Thus there exists $\mathcal{R} \in \mathcal{S}^{P^*}$, such that,

$$\mathcal{N}(\mathcal{R}) \in B_M^C(P^*), \quad \pi(\mathcal{R}) \leq \iota(M) - 1 \tag{261}$$

We then have from (258)

$$\mathcal{N}(\mathcal{R}) \leq f_0^{\pi(\mathcal{R})}(P^*) \leq f_0^{\iota(M)-1}(P^*) \tag{262}$$

which implies

$$\left\| f_0^{\iota(M)-1}(P^*) - P^* \right\| \geq \|\mathcal{N}(\mathcal{R}) - P^*\| \geq M \tag{263}$$

This contradicts the definition of $\iota(M)$ which is the smallest non-negative integer k, such that, $\left\| f_0^k(P^*) - P^* \right\| \geq M$. We thus conclude that the claim in (256) holds. ■

REFERENCES

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

[2] S. Kar, B. Sinopoli, and J. M. F. Moura, "Kalman filtering with intermittent observations: Weak convergence to a stationary distribution," March 2009, IEEE Transactions on Automatic Control (conditionally accepted as a regular paper). [Online]. Available: http://arxiv.org/abs/0903.2890

[3] H. Chen, P. Kumar, and J. van Schuppen, "On kalman filtering for conditionally gaussian systems with random matrices," *Systems & Control Letters*, vol. 13, pp. 397–404, 1989.

[4] Y. Wang and L. Guo, "On stability of random riccati equations," *Science in China Series E: Technological Sciences*, vol. 42, no. 2, pp. 136–148, April 1999.

[5] 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.

[6] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Submitted to Automatica*, 2005.

[7] Y. Xu and J. Hespanha, "Estimation under controlled and uncontrolled communications in networked control systems," in *IEEE Conference on Decision and Control*, Sevilla, Spain, December 2005, pp. 842–847.
[8] M. Huang and S. Dey, "Stability of kalman filtering with markovian packet losses," *Automatica*, vol. 43, no. 4, pp. 598–607, 2007.
[9] K. Plarre and F. Bullo, "On kalman filtering for detectable systems with intermittent observations," *IEEE Transactions on Automatic Control*, May 2007, submitted for publication.
[10] 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.
[11] L. Xie and L. Xie, "Stability of a random riccati equation with markovian binary switching," in *Proc. of IEEE CDC*, Dec 2007.
[12] 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.
[13] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME–Journal of Basic Engineering*, vol. 83, no. Series D, pp. 95–107, 1961.
[14] P. Dupuis and H. Wang, "Importance sampling, large deviations and differential games," *Stochastics and Stochastics Reports*, vol. 76, no. 6, pp. 481–508, December 2004.
[15] J. Deuschel and D. Stroock, *Large Deviations*. Academic Press, Inc., 1984.
[16] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, 2nd ed. Springer-Verlag, 1998.
[17] P. Diaconis and D. Freedman, "Iterated random functions," *SIAM REVIEW*, vol. 41, no. 1, pp. 45–76, 1999.
[18] S. Kar, B. Sinopoli, and J. M. F. Moura, "A random dynamical systems approach to filtering in large-scale networks," September 2009, to be presented at the American Control Conference, ACC 2010. [Online]. Available: http://arxiv.org/abs/0910.0918
[19] J. Jacod and A. Shiryaev, *Limit theorems for stochastic processes*. Berlin Heidelberg: Springer-Verlag, 1987.
[20] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Jones and Barlett Publishers, 1993.
[21] R. Zaharopol, *Invariant Probabilities of Markov-Feller Operators and their Supports*. Birkhauser, 2005.
[22] R. Bucy, "A priori bounds for the Riccati equation," *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, vol. 3, pp. 645–656, 1972.
[23] A. Jazwinski, *Stochastic processes and filtering theory*. Dover Publications, Inc., 1970.
[24] J. Feng and T. G. Kurtz, *Large Deviations for Stochastic Processes*. American Mathematical Society, 2006.