Balanced Realizations of Regime-Switching Linear Systems*

Y.J. Liu†   G. Yin‡   Q. Zhang§   J.B. Moore¶

March 23, 2006

Abstract

In this work, we construct a framework for balanced realization of linear systems subject to regime switching modulated by a continuous-time Markov chain with a finite state space. First, definition of balanced realization is given. Then a ρ-balanced realization is developed. When the state space of the Markov chain is large, the computational effort becomes a real concern. To resolve this problem, we introduce a two-time-scale formulation and use decomposition/aggregation and averaging techniques to reduce the computational complexity. Approximation procedures are developed. Numerical examples are also presented for demonstration.

Key words. balanced realization, continuous-time Markov chain, decomposition, aggregation, time-scale separation.

Brief Title. Balanced Realizations for Switching Systems

---

†Department of Mathematics, Missouri Southern State University, Joplin, MO 64801, liu-y@mssu.edu

‡Department of Mathematics, Wayne State University, Detroit, MI 48202, gyin@math.wayne.edu

§Department of Mathematics, The University of Georgia, Athens, GA 30602, qingz@math.uga.edu

¶Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, ACT 0200, Australia, john.moore@anu.edu.au

1   Introduction

This work is concerned with balanced realizations (or canonical realizations) of regime-switching linear systems, in which continuous dynamics and discrete events coexist. The underlying systems are modulated by a continuous-time Markov chain with state space $\mathcal{M} = \{1, \ldots, m\}$. The matrix coefficients of the linear systems depend on the states of the Markov chain. At any given instant t, the Markov chain takes one of the values from $\mathcal{M}$ (e.g., i). Then the system dynamics are determined by the matrix coefficients associated with the state i. After a random sojourn time, the Markov chain switches to a new state $j \neq i$ and stays there until the next jump; at this time the system dynamics are determined by the matrix coefficients associated with j rather than i. Such random jump linear systems arise frequently in wireless communications, manufacturing systems, financial engineering, and other applications, where regime changes are utilized to formulate the random environment. For some of the recent work on jump linear systems and their applications, we refer the reader to [10, 11, 21, 22, 23, 28] among others.

Balanced realizations have been studied for finite dimensional linear systems over the past two decades. In the 1980s, to bridge the gap between minimal realization theory and to treat the problem of finding lower order approximation, a "canonical" realization was first introduced in the seminal paper [13], where the problem was studied with principal component analysis of linear systems. The term "balance" was used since the realizations have certain symmetry between the input and the output maps characterized by the controllability and observability Grammians. Owing to its importance and its wide range of applicability, balanced realizations have attracted much attention. One of its main applications is in model reduction. While asymptotic stability of the reduced order systems was studied in [13], error bounds between the reduced order model and the original system was obtained in [6] in terms of associated singular values. There have been substantial extensions to time-varying linear systems. Main existence results concerning balanced realizations are contained in [16] and [19]. Subsequently, further work in this directions can be found in [8, 9, 14, 15] and references therein.

In this paper, we treat regime-switching linear systems or Markov jump linear systems. The systems under consideration are only piecewise deterministic, whereas the time-varying random switchings result in a much more complex structure. Our focus is on the development of "canonical realization." We keep using the term "balanced realization" although our setup is different from that of the deterministic linear systems studied in the literature. Important questions concerning such models include: How should the canonical realization be defined

in such a setting? Can the desired canonical form of these systems be reached? One of the main thoughts used in [9] is: In lieu of using exact balancing solutions in the time-varying case, we construct the so-called ρ-balancing solutions, where ρ grows monotonically to ∞ and is a natural tracking parameter. [In fact, the symbol μ in lieu of ρ is used in [9].] That is, to approximate the algebraic problems by dynamic systems with differential Riccati equations, rather than finding the exact solutions of balancing equations (time-varying algebraic equations). When regime switching is involved, we must solve not one, but a system of Riccati equations. Can error bounds of the approximation still be derived in such cases? In addition, when the number of the states of the Markov chain is large, we face large-scale systems, for which the amount of computation could be of real concern. How can we reduce the computation complexity? This paper aims to address these questions and to contribute to the following:

1. We introduce a novel model for canonical or balanced realizations of systems with regime switching. The corresponding switching mechanism regulates the moves from one configuration to another. The resulting system is piecewise deterministic.

2. We show how the balanced realization problem can be carried out for the switching models and demonstrate how systems with regime switching can be handled. Due to the time-varying feature, the balancing equation is a system of time-varying equations to be satisfied at all t. There is no feasible procedures to handle such a task. In addition, in contrast to [9], another difficulty arises since we have to deal with a coupled system of Riccati equations. As a remedy, we also use ρ-approximated balanced realizations (solutions of systems of differential Riccati equations) to approximate the system of algebraic balancing equations. Under such a setting, we derive error bounds for approximating the solution of the system of balancing equations by that of the differential system. Our numerical results show that the ρ-approximated balance preserves the robustness and the ability of coping with ill-conditioned controllability or observability Grammians.

3. We device a strategy for further reducing the computation complexity of the underlying system to treat large-scale systems. In such a setting, on top of the difficulties mentioned in the previous paragraph, the Markov chain has a large state space. It follows that solving the system of ρ-approximated balanced realization can be a computational infeasible task due to the inherent large dimensionality. To overcome the difficulties, we present a methods based on a two-fold approximation, namely, a two-time-scale

3

approximation and a ρ-approximated balance realization. We use decomposition to take advantage of the natural hierarchical structure of the systems, use an aggregation procedure to reduce the total number of states under consideration, and use averaging methods to design approximation strategy.

The rest of the paper is arranged as follows. Section 2 begins with canonical realizations when a Markov chain is the modulating force. Section 3 is concerned with reducing complexity by means of time-scale separation. Several numerical examples are presented in Section 4 for demonstration purposes. Some further remarks are made in Section 5.

2 Balanced Realizations of Regime-Switching Systems

2.1 Problem Formulation and Preliminary Results

Suppose that $\alpha(t)$ is a continuous-time Markov chain with finite state space $\mathcal{M} = \{1, 2, \ldots, m\}$ and generator

$$Q = (q_{ij}) \in \mathbb{R}^{m \times m} \ \text{ such that } \ q_{ij} \geq 0 \ \text{ for } \ i \neq j, \ \text{ and } \ \sum_{j=1}^{m} q_{ij} = 0. \tag{2.1}$$

Consider the following system

$$\frac{d}{dt}x(t) = A(\alpha(t), t)x(t) + B(\alpha(t), t)u(t), \qquad x(t_0) = x_0,$$
$$y(t) = C(\alpha(t), t)x(t), \tag{2.2}$$

where the state $x(t) \in \mathbb{R}^{n \times 1}$, input $u(t) \in \mathbb{R}^{r \times 1}$, and output $y(t) \in \mathbb{R}^{m_0}$. Note that since a finite-state Markov chain is used, effectively, (2.2) can be written as

$$\frac{d}{dt}x(t) = \sum_{i=1}^{m} A(i, t)x(t)I_{\{\alpha(t)=i\}} + \sum_{i=1}^{m} B(i, t)u(t)I_{\{\alpha(t)=i\}}, \qquad x(t_0) = x_0,$$
$$y(t) = \sum_{i=1}^{m} C(i, t)x(t)I_{\{\alpha(t)=i\}},$$

where I_S is the indicator function of the set S. In what follows, if a square matrix D is positive definite (resp. nonnegative definite), we often write it as $D > 0$ (resp. $D \geq 0$). For $D_1 \in \mathbb{R}^{\iota \times \ell}$ for some $\iota, \ell \geq 1$, D_1' denotes its transpose. Throughout the paper, for each $i \in \mathcal{M}$ and a suitable function $f(t, i)$, we denote

$$Qf(t, \cdot)(i) = \sum_{j=1}^{m} q_{ij}f(t, j) = \sum_{j \neq i} q_{ij}(f(t, j) - f(t, i)) \ \text{ for each } \ i \in \mathcal{M}.$$

To proceed, we assume condition (A1) holds.

(A1) For each $i \in \mathcal{M}$, $A(i, \cdot), B(i, \cdot), C(i, \cdot)$ are bounded and continuously differentiable functions.

Definition 2.1. For each $t \geq 0$ and $i \in \mathcal{M}$, a realization $(A(i, t), B(i, t), C(i, t))$ is said to be uniformly completely controllable if and only if there is a $\delta > 0$ such that for some positive $L_c(\delta)$ and $U_c(\delta)$,

$$\infty > U_c(\delta)I \geq G_c(t - \delta, t, i) \geq L_c(\delta)I > 0, \tag{2.3}$$

where G_c is the controllability Grammian

$$G_c(t - \delta, t, i) = \int_{t-\delta}^{t} \Phi(t, \lambda, i)B(i, \lambda)B'(i, \lambda)\Phi'(t, \lambda, i)d\lambda, \tag{2.4}$$

and $\Phi(t, \lambda, i)$ is the state transition matrix (see [2, p. 349]) of the equation $\dfrac{dz(t)}{dt} = A(i, t)z(t)$.

Definition 2.2. For each $t \geq 0$ and $i \in \mathcal{M}$, a realization $(A(i, t), B(i, t), C(i, t))$ is said to be uniformly completely observable if and only if there is a $\delta > 0$ such that for some positive $L_o(\delta)$ and $U_o(\delta)$,

$$\infty > U_o(\delta)I \geq G_o(t, t + \delta, i) \geq L_o(\delta)I > 0, \tag{2.5}$$

where G_o is the observability Grammian

$$G_o(t, t + \delta, i) = \int_{t}^{t+\delta} \Phi'(\lambda, t, i)C'(i, \lambda)C(i, \lambda)\Phi(\lambda, t, i)d\lambda. \tag{2.6}$$

As noted in [9, p. 316], if uniform complete observability and controllability hold for δ then it also holds for all $\widehat{\delta} > \delta$. Although the balancing condition depends on δ, the results to be developed does not depend on the δ selection. For notational simplicity, we suppress δ and write $G_c(t - \delta, t, i)$ as $G_c(t, i)$ and $G_o(t, t + \delta, i)$ as $G_o(t, i)$ in what follows.

Definition 2.3. The system (2.2) is said to have a balanced realization (or canonical realization) if there are nonsingular coordinate transformations $T(t, i)$ with $P(t, i) \overset{\text{def}}{=} T'(t, i)T(t, i) > 0$ such that

$$P(t, i)G_c(t, i)P(t, i) = G_o(t, i) + QP(t, \cdot)(i), \qquad i = 1, 2, \ldots, m. \tag{2.7}$$

Our objective is to find balanced realizations of the regime-switching linear systems. To simplify the notation, we write (2.7) in a matrix form. To proceed, introduce the following notation:

$$\begin{aligned}
P(t) &= \text{diag}(P(t, 1), \ldots, P(t, m)), \\
G_c(t) &= \text{diag}(G_c(t, 1), \ldots, G_c(t, m)), \\
G_o(t) &= \text{diag}(G_o(t, 1), \ldots, G_o(t, m)),
\end{aligned} \tag{2.8}$$

where diag($\cdot$) denotes a diagonal block matrix with appropriate dimensions. Note that in view of the notation above, $P(t) = \text{diag}(P(t,1), \ldots, P(t,m)) > 0$ since $P(t,i) > 0$ for each $i \in \mathcal{M}$. Thus, we can rewrite the system of algebraic Riccati equations (2.7) as

$$P(t)G_c(t)P(t) = G_o(t) + \text{diag}\Big(\sum_{j=1}^{m} q_{1j}P(t,j), \ldots, \sum_{j=1}^{m} q_{mj}P(t,j)\Big). \qquad (2.9)$$

A regime-switching linear system has a balanced realization if (2.9) holds. By a first glance, (2.9) can be solved for each fixed t. As a result, the problem seems to be largely resolved. In fact, it barely begins. The reason is that we need to solve (2.9) for all t not just a fixed t or finitely many t. Directly solving this system is not feasible for all t. Besides, compared to the deterministic counter part, it is a coupled system of equations, which adds another fold of difficulty. To overcome the difficulties, we approximate the algebraic system by a system of differential equations.

To proceed, we rewrite the last term of (2.9) in terms of $P(t)$. Denoting by I_n the $n \times n$ identity matrix, $\widehat{I}_{mn} = (I_n, \ldots, I_n)' \in \mathbb{R}^{(mn) \times n}$, and using $Q \otimes I_n$, the Kronecker product [7, p. 21], we have

$$Q \otimes I_n = \begin{pmatrix} q_{11}I_n & q_{12}I_n & \cdots & q_{1m}I_n \\ q_{21}I_n & q_{22}I_n & \cdots & q_{2m}I_n \\ & \cdots & \cdots & \\ q_{m1}I_n & q_{m2}I_n & \cdots & q_{mm}I_n \end{pmatrix} \in \mathbb{R}^{(mn) \times (mn)},$$

and

$$(Q \otimes I_n)P(t)\widehat{I}_{mn} = \Big(\sum_{j=1}^{m} q_{1j}P(t,j), \ldots, \sum_{j=1}^{m} q_{mj}P(t,j)\Big)'$$
$$\stackrel{\text{def}}{=} P_{mn}(t) \in \mathbb{R}^{(mn) \times n}.$$

To obtain the last term in (2.9), note that

$$(Q \otimes I_n)P(t)\widehat{I}_{mn}I_0 = \begin{pmatrix} \sum_{j=1}^{m} q_{1j}P(t,j) & 0_n & \cdots & 0_n \\ \sum_{j=1}^{m} q_{2j}P(t,j) & 0_n & \cdots & 0_n \\ & \cdots & & \cdots \cdots \cdots \\ \sum_{j=1}^{m} q_{mj}P(t,j) & 0_n & \cdots & 0_n \end{pmatrix} \in \mathbb{R}^{(mn) \times (mn)},$$

where 0_n is an $n \times n$ zero matrix, and $I_0 = (I_n, 0_n, \ldots, 0_n)$ is an $n \times (mn)$ matrix. Let E_{ij} be a block matrix that has dimension $(mn) \times (mn)$ with the (i,j) element being I_n and all other elements being 0_n. The reason for introducing E_{ij} is that it enables us to write the

equations of interest in a more manageable form and to carry out desired calculations more easily. Then we have

$$
E_{ii}[(Q \otimes I_n)P(t)(\widehat{I}_{mn}I_0)]E_{1i} =
\begin{pmatrix}
0_n & 0_n & \cdots & & 0_n \\
0_n & 0_n & & \cdots & 0_n \\
\cdots & \cdots & \sum_{j=1}^m q_{ij}P(t,j) & \cdots & \\
\cdots & \cdots & \cdots & & \cdots \\
0_n & \cdots & 0_n & & 0
\end{pmatrix}
\in \mathbb{R}^{(mn)\times(mn)},
$$

and hence

$$
\sum_{i=1}^m [E_{ii}(Q \otimes I_n)]P(t)[(\widehat{I}_{mn}I_0)E_{1i}] = \mathrm{diag}\Big(\sum_{j=1}^m q_{1j}P(t,j), \ldots, \sum_{j=1}^m q_{mj}P(t,j) \Big).
$$

Therefore, (2.9) can be written as

$$
P(t)G_c(t)P(t) = G_o(t) + \sum_{i=1}^m [E_{ii}(Q \otimes I_n)]P(t)[(\widehat{I}_{mn}I_0)E_{1i}]. \tag{2.10}
$$

Write

$$
G_i = E_{ii}(Q \otimes I_n), \quad H_i = \widehat{I}_{mn}I_0 E_{1i}.
$$

The use of E_{ij} enables us conveniently to rewrite equation (2.10) as

$$
P(t)G_c(t)P(t) = G_o(t) + \sum_{i=1}^m G_i P(t) H_i. \tag{2.11}
$$

Use $\lambda(\widehat{A})$ to denote the eigenvalue of the matrix $\widehat{A}$ and use the induced matrix norm $|\widehat{A}| = \sup_{|x|=1} |\widehat{A}x| = \sqrt{\lambda_{\max}(\widehat{A}'\widehat{A})}$. We need the following assumption.

(A2) For $i \in \mathcal{M}$, $(A(i,t), B(i,t), C(i,t))$ is uniformly completely controllable and observable.

Following the idea in [9], introduce the system of Riccati equations to approximate the exact solutions of (2.9)

$$
\frac{d}{d\tau}P_\rho(\tau : t, t_0, i) = -\rho P_\rho(\tau : t, t_0, i)G_c(t,i)P_\rho(\tau : t, t_0, i) + \rho G_o(t,i) + \rho Q P(\tau : t, t_0, \cdot)(i),
$$
$$
P_\rho(t_0 : t, t_0, i) \quad > 0, \text{ for } i \in \mathcal{M},
$$
$$
\tag{2.12}
$$

where $\rho > 0$ is an arbitrary constant gain. The rationale here is to constructing ρ-balancing, or a ρ-approximated canonical realization, where ρ indicates the degree of approximation of balancing.

Under conditions (A1) and (A2), similar to [27], we can prove that there exists a unique positive definite solution $P_\rho(\tau : t, t_0, i)$. Write $P_\rho(t : t, t_0, i) = P_\rho(t, t_0, i)$, we now consider the following system of Riccati differential equations parameterized by $\rho > 0$,

$$\frac{d}{dt} P_\rho(t, t_0, i) = -\rho P_\rho(t, t_0, i) G_c(t, i) P_\rho(t, t_0, i) + \rho G_o(t, i) + \rho Q P(t, t_0, \cdot)(i),$$
$$P_\rho(t_0, t_0, i) > 0, \text{ for } i \in \mathcal{M}. \tag{2.13}$$

It can be shown that the limit

$$\widehat{P}_\rho(t, i) \stackrel{\text{def}}{=} \lim_{t_0 \to -\infty} P_\rho(t, t_0, i) \tag{2.14}$$

exists and satisfies the following system of Riccati differential equations

$$\frac{d}{dt} \widehat{P}_\rho(t, i) = -\rho \widehat{P}_\rho(t, i) G_c(t, i) \widehat{P}_\rho(t, i) + \rho G_o(t, i) + \rho Q \widehat{P}(t, \cdot)(i), \ i \in \mathcal{M}. \tag{2.15}$$

Definition 2.4. Given a realization $(A(i, t), B(i, t), C(i, t) : i \in \mathcal{M})$ with Grammians $(G_c(t, i),$ $G_o(t, i) : i \in \mathcal{M})$, the transformation $(T_\rho(t, i) : i \in \mathcal{M})$ is ρ-balancing if $\widehat{P}(t, i) = T_\rho(t, i) T_\rho(t, i)$ satisfies (2.14) for all $t \in (-\infty, \infty)$ and $i \in \mathcal{M}$.

To proceed, we state a lemma, whose proofs is relegated to an appendix in order to keep the flow of presentation.

Lemma 2.5. *Consider* (2.2) *and suppose conditions* (A1) *and* (A2) *hold. Let* $P(t)$ *be the solution of* (2.11). *Then there exist some constants* $\gamma_\iota > 0$ *(for* $\iota = 1, \ldots, 5$) *such that for* $t \in [0, \infty)$,

$$\gamma_2 I \geq P(t) \geq \gamma_1 I,$$
$$\lambda_{\min}(P(t) G_c(t)) \geq \gamma_3,$$
$$\lambda_{\max}(P(t) G_c(t)) \leq \gamma_4, \tag{2.16}$$
$$-\gamma_5 I \leq \dot{P}(t) \leq \gamma_5 I.$$

2.2 Error Bounds of ρ-Balanced Realizations

Although (2.9) provides an exact system of balancing equations, since it is time-varying, the solution must be obtained for all t. Although it is possible to solve the system at some fixed time t, there is virtually no feasible procedure that can carry out the desired task for all t. As a remedy, in light of [9], we approximate the exact solution of the balancing equation (2.9) by its dynamic system counter part. Compared with [9], another difficulty arises since we have to deal with a coupled system of Riccati equations. The main result here is on the approximation error bounds, which are presented in the next two theorems.

Theorem 2.6. *Suppose conditions* (A1) *and* (A2) *hold. Denote*

$$P_\rho(\tau : t, t_0) = \mathrm{diag}(P_\rho(\tau : t, t_0, 1), \dots, P_\rho(\tau : t, t_0, m)),$$

$$\triangle P_\rho(\tau : t, t_0) = P_\rho(\tau : t, t_0) - P(t),$$

where $P(t)$ is defined in (2.8). *Then for any $\rho > 0$,*

$$| \triangle P_\rho(\tau : t, t_0)| \leq | \triangle P_\rho(t_0 : t, t_0)| e^{-\rho \widetilde{\gamma}_3 (t - t_0)}, \quad for \ \ t, \tau \in [t_0, \infty). \tag{2.17}$$

Proof. Using matrix form (2.12) and (2.11), we have

$$\frac{d}{d\tau}[P_\rho(\tau : t, t_0) - P(t)] = \rho \Big[- P(t) G_c(t)(P_\rho(\tau : t, t_0) - P(t))$$
$$- (P_\rho(\tau : t, t_0) - P(t)) G_c(t) P_\rho(\tau : t, t_0)$$
$$+ \sum_{i=1}^{m} G_i((P_\rho(\tau : t, t_0) - P(t))) H_i \Big].$$

Thus, we obtain

$$P_\rho(\tau : t, t_0) - P(t)$$
$$= \exp(-\rho(P(t) G_c(t))(\tau - t_0))(P_\rho(t_0 : t, t_0) - P(t))$$
$$- \rho \int_{t_0}^{\tau} \exp(-\rho(P(t) G_c(t))(s - t_0))(P_\rho(s : t, t_0) - P(t)) G_c(t) P_\rho(s : t, t_0) ds$$
$$+ \rho \int_{t_0}^{\tau} \exp(-\rho(P(t) G_c(t))(s - t_0)) \sum_{i=1}^{m} G_i((P_\rho(s : t, t_0) - P(t))) H_i ds.$$

Noting the positive definiteness of $G_c(\cdot)$ and $P_\rho(\tau : t, t_0)$ and by virtue of (2.16), we have

$$\lambda_{\min}(P_\rho(\tau : t, t_0) G_c(t)) \geq \lambda_{\min}(P_\rho(\tau : t, t_0) \lambda_{\min}(G_c(t)) = \lambda_{\min}(P_\rho(\tau : t, t_0) L_c := \widehat{\gamma}_3.$$

Therefore,

$$| P_\rho(\tau : t, t_0) - P(t)|$$
$$\leq | \exp(-\rho(P(t) G_c(t))(\tau - t_0))||(P_\rho(t_0 : t, t_0) - P(t))| \tag{2.18}$$
$$+ K\rho \int_{t_0}^{\tau} | \exp(-\rho(P(t) G_c(t))(s - t_0))||(P_\rho(s : t, t_0) - P(t))| ds,$$

where

$$K = \widehat{\gamma}_3 + \sum_{i=1}^{m} |G_i||H_i|.$$

9

By virtue of the Gronwall inequality, we obtain

$$
\begin{aligned}
|P_\rho(\tau : t, t_0) - P(t)| \leq{}& K\rho |\exp(-\rho(P(t)G_c(t))(\tau - t_0))||(P_\rho(t_0 : t, t_0) - P(t))| \\
& \times \exp(\int_{t_0}^{\tau} |\exp(-\rho(P(t)G_c(t))(s - t_0))|ds) \\
\leq{}& K\rho |\exp(-\rho(P(t)G_c(t))(\tau - t_0))||(P_\rho(t_0 : t, t_0) - P(t))| \quad (2.19) \\
\leq{}& K\rho \exp(-\lambda_3 \rho(\tau - t_0))|(P_\rho(t_0 : t, t_0) - P(t))| \\
\leq{}& \exp(-\widetilde{\lambda}_3 \rho(\tau - t_0))|(P_\rho(t_0 : t, t_0) - P(t))|,
\end{aligned}
$$

where $0 < \widetilde{\lambda}_3 < \lambda_3$. $\quad\square$

Theorem 2.7. *Suppose conditions* (A1) *and* (A2) *hold. Then* $\Delta P_\rho(t, t_0) = P_\rho(t, t_0) - P(t)$ *is bounded by*

$$
|\Delta P_\rho(t, t_0)| \leq |\Delta P_\rho(t_0, t_0)| e^{-\rho \widetilde{\gamma}_3 (t - t_0)} + \frac{\gamma_5}{\rho \widetilde{\gamma}_3}, \qquad for\ t \in [t_0, \infty) \qquad (2.20)
$$

Moreover, $\Delta \widehat{P}_\rho(t) = \widehat{P}_\rho(t) - P(t)$ *has the following properties*

$$
\begin{aligned}
|\Delta \widehat{P}_\rho(t)| \quad &\leq \frac{\gamma_5}{\rho \widetilde{\gamma}_3}, \ \ for\ t \in (-\infty, \infty), \\
\lim_{\rho \to \infty} \Delta \widehat{P}_\rho(t) &= 0 \qquad uniformly\ for\ t\ \ in\ any\ compact\ interval.
\end{aligned}
\qquad (2.21)
$$

Proof. To facilitate the proof, we first use a piecewise constant approximation to the Grammian, which will enable us to derive the desired error bound relatively easily. Note that these piecewise constant functions only serve auxiliary machinery; they are not used in the actual realizations.

To this end, let the Grammians a constant on a small interval (t_i, t_i+1) with $t_{i+1} - t_i = \eta$ and $\eta > 0$ is sufficiently small. We denote them with a tilde and require $\widetilde{G}_c(t_i) = G_c(t_i)$, $\widetilde{G}_o(t_i) = G_o(t_i)$. By virtue of (2.16) the maximal change over any discontinuity is $\gamma_5 \eta$. Then we obtain

$$
\begin{aligned}
|\Delta P_\rho(t_1, t_0)| &= |\widetilde{P}_\rho(t_1, t_0) - \widetilde{P}(t_1)| \\
&\leq |\widetilde{P}_\rho(t_1, t_0) - \widetilde{P}(t_0)| + |\widetilde{P}(t_0) - \widetilde{P}(t_1)| \\
&\leq e^{-\rho \widetilde{\gamma}_3 \eta} |\Delta P_\rho(t_0, t_0)| + \gamma_5 \eta.
\end{aligned}
$$

10

Similarly, we have

$$
\begin{aligned}
|\Delta P_\rho(t_i, t_0)| = |\widetilde{P}_\rho(t_i, t_0) - \widetilde{P}(t_i)| \\
\leq e^{-\rho\widetilde{\gamma}_3\eta}|\Delta P_\rho(t_{i-1}, t_0)| + \gamma_5\eta \\
\leq e^{-i\rho\widetilde{\gamma}_3\eta}|\Delta P_\rho(t_0, t_0)| + \sum_{j=0}^{i-1} e^{-j\rho\widetilde{\gamma}_3\eta}\gamma_5\eta \\
\leq e^{-i\rho\widetilde{\gamma}_3\eta}|\Delta P_\rho(t_0, t_0)| + \sum_{j=0}^{\infty} e^{-j\rho\widetilde{\gamma}_3\eta}\gamma_5\eta \\
\leq e^{-i\rho\widetilde{\gamma}_3(t-t_0)}|\Delta P_\rho(t_0, t_0)| + \frac{\gamma_5\eta}{1 - e^{-\rho\widetilde{\gamma}_3\eta}}.
\end{aligned}
$$

Next, we consider the above estimate on the interval $[t_i, t_{i+1})$,

$$
\begin{aligned}
|\Delta P_\rho(t, t_0)| = |\widetilde{P}_\rho(t, t_0) - \widetilde{P}(t_i)| \\
\leq e^{-\rho\widetilde{\gamma}_3(t-t_i)}|\widetilde{P}_\rho(t_i, t_0) - \widetilde{P}(t_i)| \\
\leq e^{-\rho\widetilde{\gamma}_3(t-t_0)}|\Delta P_\rho(t_0, t_0)| + \frac{\gamma_5\eta}{1 - e^{-\rho\widetilde{\gamma}_3\eta}}.
\end{aligned}
\tag{2.22}
$$

For piecewise constant Grammians with possible jumps at t_i, and $t \in [t_i, t_{i+1})$,

$$
|\widetilde{P}_\rho(t, t_0) - P(t)| \leq e^{-\rho\widetilde{\gamma}_3(t-t_0)}|\Delta P_\rho(t_0, t_0)| + \frac{\gamma_5\eta}{1 - e^{-\rho\widetilde{\gamma}_3\eta}}.
$$

Since $P(t)$ and $P_\rho(t, t_0)$ are continuous, and piecewise constant functions are dense in the class of continuous functions, using standard approximation arguments we can show that $\widetilde{P}_\rho(t_i, t_0) \to \widetilde{P}(t)$ and $P(t_i) \to P(t)$ as $t_{i+1} - t_i \to 0$. Finally, we obtain

$$
\begin{aligned}
|P_\rho(t, t_0) - P(t)| = \lim_{\eta \to 0} |\widetilde{P}_\rho(t_i, t_0) - \widetilde{P}(t_i)| \\
\leq |\Delta P_\rho(t_0, t_0)| e^{-\rho\widetilde{\gamma}_3(t-t_0)} + \frac{\gamma_5}{\rho\widetilde{\gamma}_3} \qquad \text{for all } t \in (t_0, \infty).
\end{aligned}
$$

Sending $t_0 \to -\infty$, we obtain the first part of (2.21). Letting $\rho \to \infty$ we obtain the second part of (2.21). $\qquad \square$

3 Large-scale Systems: Further Reduction of Complexity

This section is concerned with balanced realizations for regime-switching systems when the underlying Markov chain has a large state space. Even if we construct ρ-balancing systems of equations, the large-scale nature of the systems can render the computation task infeasible

11

due to the large dimensionality. Alternative methods must be designed. We suggest to treat the problem by using a two-fold approximation. That is, in addition to the ρ-balancing, we introduce a two-time-scale model, and examine two-time-scale ρ-balancing systems of equations. The two-time scale is represented using a small parameter $\varepsilon > 0$. Using inherent fast and slow time scales, first let $\varepsilon \to 0$, we obtain a limit ρ-balancing system, and then we can use the ρ-balancing to approximate (2.9) as in the last section.

3.1   Two-time-scale Setup

This section focuses on reduction of computational complexity. As observed in [18], there are certain hierarchical structures in every large-scale and complex system. Not all the components or subsystems in a large-scale model vary at the same rate. Some of them change rapidly, whereas others evolve slowly. It is advantageous to bring out the hierarchical structure. One way of doing this is to introduce a model involving multiple time scales. Studying applications in queueing and computer systems, the so-called nearly completely decomposable models came into being (see [4]). The main idea lies in the utilization of time-scale separation, which enables us to decompose the underlying system into subsystems and to aggregate the states in such subsystems into one state, leading to substantial savings in terms of computational effort. In this section, we suppose that the generator of the Markov chain has the following form

$$Q^\varepsilon = \frac{1}{\varepsilon}\widetilde{Q} + \widehat{Q} = (q_{ij}^\varepsilon), \tag{3.1}$$

where both $\widetilde{Q}$ and $\widehat{Q}$ are themselves generators of continuous-time Markov chains, and $\varepsilon > 0$ is a small parameter. The matrix $\widetilde{Q}$ represents the rapidly changing part and $\widehat{Q}$ describes the slowly varying part. A reader may consult [23, Section 3.6] to see how a practical system may be reduced to the form (3.1). The rationale is that the Markov chain is almost decomposable into l ergodic classes of states. Nevertheless, due to the weak interactions resulted from the generator $\widehat{Q}$, there are still infrequent transitions from one ergodic class $\mathcal{M}_i$ to another ergodic class $\mathcal{M}_j$ for $i \neq j$. In a way, the state space of the Markov chain is only nearly separated into l ergodic classes. This is more realistic than the formulation of completely separable classes. For convenience, in what follows, we relabel the states so that $\mathcal{M}_k = \{s_{k1}, \ldots, s_{km_k}\}$, and the state space of Markov chain $\alpha(\cdot)$ is $\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_l$.

Now, we have the similar equations as in (2.11) and (2.13) indexed by ε. Suppose the rapidly changing generator $\widetilde{Q}$ has the form $\widetilde{Q} = \text{diag}(\widetilde{Q}^1, \ldots, \widetilde{Q}^l)$ where $\widetilde{Q}^k \in \mathbb{R}^{m_k \times m_k}$ are irreducible, for $k = 1, 2, \ldots, l$ and $\sum_{k=1}^l m_k = m$. The state spaces of $\widetilde{Q}^k$ is $\mathcal{M}_k$. By aggregating the states in $\mathcal{M}_k$ as one state k, we obtain an aggregated process $\{\overline{\alpha}^\varepsilon(\cdot)\}$, i.e.,

$\overline{\alpha}^\varepsilon(t) = k$ if $\alpha(t) \in \mathcal{M}_k$. It is shown in [23] that the process $\overline{\alpha}^\varepsilon(t)$ is not necessarily Markovian but $\overline{\alpha}^\varepsilon(\cdot)$ converges weakly to $\overline{\alpha}(\cdot)$, a Markov chain with generator

$$\overline{Q} = \mathrm{diag}(\nu^1, \ldots, \nu^l)\widehat{Q}\mathrm{diag}(\mathbb{1}_{m_1}, \ldots, \mathbb{1}_{m_l}), \tag{3.2}$$

where ν^k is the stationary distribution of $\widetilde{Q}^k$, for $k = 1, \ldots, l$, and $\mathbb{1}_\iota = (1, \ldots, 1)' \in \mathbb{R}^\iota$. Moreover, for any bounded deterministic $\beta(\cdot)$,

$$\sup_{0 \le t \le T} E \left(\int_0^t [I_{\{\alpha^\varepsilon(s)=s_{kj}\}} - \nu_j^k I_{\{\overline{\alpha}^\varepsilon(s)=k\}}]\beta(s)ds \right)^2 = O(\varepsilon). \tag{3.3}$$

If the original Markov chain has m states and the limit of the aggregated process has l states, and if $l \ll m$, then substantial reduction of complexity is achieved.

Remark 3.1. Before proceeding further, to put this work in perspective, we compare our results with that in the literature. Our formulation is motivated by the nearly completely decomposable matrix models of [4, 18]. It has the same goal–reduction of complexity. Nevertheless, there are certain differences. (a) The models considered in [4, 18] are in discrete time. In these references,

$$P^\varepsilon = P + \varepsilon C, \tag{3.4}$$

where P is a transition matrix that consists of a number of irreducible transition matrices and C, containing some negative entries, is a matrix that together with P makes P^ε a transition matrix. In contrast, our formulation is in continuous time. Although $\widetilde{Q}$ has similar decomposition form as that of [4, 18], $\widetilde{Q}$ and $\widehat{Q}$ are generators (with row sums being 0) instead of transition matrices (with row sums being 1). In lieu of small perturbations, we consider large perturbations. (b) Our model puts more emphases on the probabilistic structure. Note that both $\widetilde{Q}$ and $\widehat{Q}$ are generators, which facilitate the analysis as well as interpretation of the results. (c) The reference [4] presents procedures for computing stationary distribution or steady state probabilities. Computing the overall stationary distribution of the large Markov chain is not our main concern. We aim at bring out the dynamic structure of the limit system. For a Markov chain having a large state space (with m elements and m being a large number), using the results of last section, it is readily seen that the ρ-approximate realization results in a system of differential equations. The total number of the equations is m. By using a two-time-scale approach, we obtain a limit or reduced order system, which has only l equations. When $l \ll m$, a substantial reduction of computation effort is achieved. In a nutshell, the coefficients of the limit or reduced order system are that of the original system averaged out with respect to the stationary measures associated with the $\widetilde{Q}^i$'s for

$i = 1, \ldots, l$. (d) Our recent work [24] examines discrete-time model as in (3.4). However, in our setup C is a generator of a continuous-time Markov chain. This enables us to precisely describe the associated limit or reduced order systems.

To proceed, for any $k = 1, \ldots, l$, and any function on $\mathcal{M}_k$, we define $\overline{F}(k) = \sum_{j=1}^{m_k} \nu_j^k F(s_{kj})$. To proceed, we device a two-stage approximation scheme. First, for a fixed ρ, we show that a system of Riccati equations involving a small parameter ε converges to a limit system as $\varepsilon \to 0$. Then we use the results obtained in the last section to link the original system and the balanced system. Similar to the last section, with ε used, define $P^\varepsilon(t, s_{kj})$ analogous to $P(t, i)$ in the last section, and define $P_\rho^\varepsilon(t, t_0, s_{kj})$ analogous to $P_\rho(t, t_0, i)$ in the last section.

Theorem 3.2. *For $k = 1, \ldots, l$ and $j = 1, \ldots, m_k$,*

(i) *$P_\rho^\varepsilon(t, t_0, s_{kj}) \to \overline{P}_\rho(t, t_0, k)$ uniformly on $[t_0, T]$ where $\overline{P}_\rho(t, t_0, k)$ is the unique solution of the following system of Riccati differential equations*

$$\frac{d}{dt} \overline{P}_\rho(t, t_0, k) = -\rho \overline{P}_\rho(t, t_0, k) \overline{G}_c(t, k) \overline{P}_\rho(t, t_0, k) + \rho \overline{G}_o(t, k) + \rho \overline{Q} \ \overline{P}_\rho(t, t_0, \cdot)(k),$$
(3.5)

(ii) *$P^\varepsilon(t, s_{kj}) \to \overline{P}(t, k)$, where $\overline{P}(t, k)$ satisfies*

$$\overline{P}(t, k) \overline{G}_c(t, k) \overline{P}(t, k) = \overline{G}_o(t, k) + \overline{Q} \ \overline{P}(t, \cdot)(k),$$
(3.6)

with

$$\overline{G}_c(t, k) = \sum_{j=1}^{m_k} \nu_j^k G_c(t, s_{kj}) \quad and \quad \overline{G}_o(t, k) = \sum_{j=1}^{m_k} \nu_j^k G_o(t, s_{kj}),$$

and $\overline{Q}$ being given by (3.2).

Remark 3.3. It is easily seen that the total number of equations to be solved in both the balancing equation and dynamic Riccati equations are substantially reduced if $l \ll m$. In what follows, we will verify only (i) in the above theorem and obtain the limit Riccati equations. The algebraic equations (ii) can be obtained in a similar fashion with simpler argument. As a preparation, we first state several lemmas.

Lemma 3.4. *The systems of Riccati equations (2.13) with Q replaced by Q^ε admit a unique solution $P_\rho^\varepsilon(t, t_0, i)$. Moreover, the solution is positive definite for $i \in \mathcal{M}$ and $t_0 \le t \le T$.*

Proof. The proof is similar to [27], and we omit the details. □

Lemma 3.5. *For fixed ρ, the solution $P_\rho^\varepsilon(t, t_0, i)$ is uniformly bounded in ε and $t \in [t_0, T]$.*

Proof. Rewriting (2.13) in vector form, we have

$$\frac{d}{dt} P_\rho^\varepsilon(t, t_0) = \rho Q^\varepsilon P_\rho^\varepsilon(t, t_0) + \rho G_o(t) - \rho P_\rho^{\varepsilon,\prime}(t, t_0) G_0(t) P_\rho^\varepsilon(t, t_0), \tag{3.7}$$

where

$$P_\rho^\varepsilon(t, t_0) = (P_\rho^\varepsilon(t, t_0, 1), \ldots, P_\rho^\varepsilon(t, t_0, m))', \quad G_o(t) = (G_o(t, 1), \ldots, G_o(t, m))'.$$

Introducing a new variable $\tau = t/\varepsilon$, then $\tau_0 = t_0/\varepsilon$ and (3.7) can be written as

$$\frac{d}{d\tau} P_\rho^\varepsilon(\tau, \tau_0) = \rho Q^\varepsilon P_\rho^\varepsilon(\tau, \tau_0) + \rho G_0(\varepsilon\tau) - \rho P_\rho^{\varepsilon,\prime}(\tau, \tau_0) G_0(\varepsilon\tau) P_\rho^\varepsilon(\tau, \tau_0).$$

It follows that

$$\begin{aligned}
P^\varepsilon(\tau, \tau_0) =\ & P_\rho^\varepsilon(\tau_0, \tau_0) \exp(\rho \widetilde{Q} \tau) \\
& - \rho \int_{\tau_0}^\tau \exp(\rho \widetilde{Q}(\tau - s)) P_\rho^{\varepsilon,\prime}(s, \tau_0) G_0(\varepsilon s) P_\rho^\varepsilon(s, \tau_0) ds \\
& + \rho \int_{\tau_0}^\tau \exp(\rho \widetilde{Q}(\tau - s)) G_0(\varepsilon s) ds \\
& + \varepsilon \rho \int_{\tau_0}^\tau \exp(\rho \widetilde{Q}(\tau - s)) \widehat{Q} P_\rho^\varepsilon(s, \tau_0) ds.
\end{aligned} \tag{3.8}$$

Recall that we use K to represent a generic positive constant; its value may change for different usages. Note that

$$\begin{aligned}
& \left| \int_{\tau_0}^\tau \exp(\rho \widetilde{Q}(\tau - s)) G_0(\varepsilon s) ds \right| \\
& \leq \left| \int_{\tau_0}^\tau [\exp(\rho \widetilde{Q}(\tau - s)) - \rho\nu] G_0(\varepsilon s) ds \right| + \rho \left| \nu \int_{\tau_0}^\tau G_0(\varepsilon s) ds \right| \\
& \leq K \int_{\tau_0}^\tau \exp(-\rho\kappa_0(\tau - s)) |G_0(\varepsilon s)| ds + K\rho \int_{\tau_0}^\tau |G_0(\varepsilon s)| ds \\
& \leq K + K\rho \int_{t_0}^t |G_0(s)| ds \leq K.
\end{aligned} \tag{3.9}$$

Note also that

$$\left| P_\rho^\varepsilon(\tau_0, \tau_0) \exp(\rho \widetilde{Q} \tau) \right| \leq \left| P_\rho^\varepsilon(\tau_0, \tau_0) [\exp(\rho \widetilde{Q} \tau) - \rho\nu] \right| + |P_\rho^\varepsilon(\tau_0, \tau_0) \rho\nu| \leq K, \tag{3.10}$$

and that

$$\left| \rho\varepsilon \int_{\tau_0}^\tau \rho\nu \widehat{Q} ds \right| \leq K\rho^2 \varepsilon \frac{t - t_0}{\varepsilon} \leq K\rho^2 T. \tag{3.11}$$

Moreover, taking trace in (3.8), we arrive at

$$\begin{aligned}
\mathrm{tr}P^{\varepsilon}(\tau,\tau_0) \leq\ & \mathrm{tr}[P^{\varepsilon}_{\rho}(\tau_0,\tau_0)\exp(\rho\widetilde{Q}\tau)] \\
& +\rho\mathrm{tr}\left[\int_{\tau_0}^{\tau}\exp(\rho\widetilde{Q}(\tau-s))G_0(\varepsilon s)ds\right] \\
& +\varepsilon\rho\mathrm{tr}\left[\int_{\tau_0}^{\tau}\exp(\rho\widetilde{Q}(\tau-s))\widehat{Q}P^{\varepsilon}_{\rho}(s,\tau_0)ds\right].
\end{aligned} \tag{3.12}$$

Using (3.9)–(3.11), an application of the Gronwall's inequality leads to

$$|\mathrm{tr}P^{\varepsilon}_{\rho}(\tau,\tau_0)| \leq K\exp\left(\varepsilon\rho\int_{\tau_0}^{\tau}ds\right) \leq K,$$

where K is a generic positive constant that may depend on ρ and T but is independent of ε. The desired result follows. $\quad\square$

Lemma 3.6. $P^{\varepsilon}_{\rho}(t,t_0,i)$ is uniformly Lipschitz continuous on $[t_0,T]$ for $i \in \mathcal{M}$.

Proof. By multiplying both sides of (3.7) by the stationary distribution $\nu = (\nu^1,\ldots,\nu^l)$,

$$\nu Q^{\varepsilon} = \nu\left(\frac{\widetilde{Q}}{\varepsilon}+\widehat{Q}\right) = \nu\widehat{Q},$$

and rewriting (3.7) in integral form, we have

$$P^{\varepsilon}_{\rho}(t,t_0) - P^{\varepsilon}_{\rho}(s,t_0) = \int_{s}^{t}[-\rho P^{\varepsilon,\prime}_{\rho}(r,t_0)G_o(r)P^{\varepsilon}_{\rho}(r,t_0) + \rho G_o(r) + \rho Q^{\varepsilon}P^{\varepsilon}_{\rho}(r,t_0)]dr. \tag{3.13}$$

Multiplying the above equation by $\nu = (\nu^1,\ldots,\nu^l)$, we have

$$\nu(P^{\varepsilon}_{\rho}(t,t_0) - P^{\varepsilon}_{\rho}(s,t_0)) = \int_{s}^{t}[-\rho\nu P^{\varepsilon,\prime}_{\rho}(r,t_0)G_o(r)P^{\varepsilon}_{\rho}(r,t_0) + \rho\nu G_o(r) + \rho\nu\widehat{Q}P^{\varepsilon}_{\rho}(r,t_0)]dr. \tag{3.14}$$

By virtue of the uniform boundedness of P^{ε}_{ρ}, G_c, and G_o, we obtain

$$\begin{aligned}
|\nu(P^{\varepsilon}_{\rho}(t,t_0) - P^{\varepsilon}_{\rho}(s,t_0))| \leq\ & \int_{s}^{t}\left[\left|-\rho\nu P^{\varepsilon,\prime}_{\rho}(r,t_0)G_o(r)P^{\varepsilon}_{\rho}(r,t_0)\right| + \left|\rho\nu G_o(r)\right|\right. \\
& \left. +\left|\rho\nu\widehat{Q}P^{\varepsilon}_{\rho}(r,t_0)\right|\right]dr \\
\leq\ & K|t-s|.
\end{aligned} \tag{3.15}$$

Then it follows

$$\begin{aligned}
|P^{\varepsilon}_{\rho}(t,t_0,i) - P^{\varepsilon}_{\rho}(s,t_0,i)| &\leq K|P^{\varepsilon}_{\rho}(t,t_0) - P^{\varepsilon}_{\rho}(s,t_0)| \\
&\leq K|\nu(P^{\varepsilon}_{\rho}(t,t_0) - P^{\varepsilon}_{\rho}(s,t_0))| \\
&\leq K|t-s|.
\end{aligned} \tag{3.16}$$

The proof is completed. $\square$

Proof of Theorem 3.2. By virtue of Lemmas 3.5 and 3.6, for any $i = s_{kj} \in \mathcal{M}_k$, $P_\rho^\varepsilon(t, t_0, s_{kj})$ is uniformly bounded and equicontinuous. By the Arzela-Ascoli Theorem, for any sequence indexed by ε, we can extract a subsequence still indexed by ε (for notational simplicity) such that $P_\rho^\varepsilon(t, t_0, s_{kj})$ converges uniformly on $[t_0, T]$ to a continuous function denoted by $P_\rho^0(t, t_0, s_{kj})$. First, we claim that $P_\rho^0(t, t_0, s_{kj})$ is independent of j. Rewriting the Riccati equations in the integral form, we obtain

$$P_\rho^\varepsilon(t, t_0, s_{kj}) = P_\rho^\varepsilon(t_0, t_0, s_{kj}) + \int_{t_0}^t [-\rho P_\rho^\varepsilon(r, t_0, s_{kj}) G_c(r, s_{kj}) P_\rho^\varepsilon(r, t_0, s_{kj}) + \rho G_o(r, s_{kj})$$
$$+ \rho Q^\varepsilon P_\rho^\varepsilon(r, t_0, \cdot)(s_{kj})] dr.$$
(3.17)

Since Q^ε is given by (3.1) and $P_\rho^\varepsilon(\cdot)$ is uniformly bounded, we multiply both sides of (3.17) by ε to obtain

$$\lim_{\varepsilon \to 0} \int_{t_0}^t \widetilde{Q}^k P_\rho^\varepsilon(r, t_0, \cdot)(s_{kj}) dr = \int_{t_0}^t \widetilde{Q}^k P_\rho^0(r, t_0, \cdot)(s_{kj}) dr = 0, \quad \text{for } t \in [t_0, T].$$

Since $P_\rho^0(t, t_0, s_{kj})$ is continuous in t, we obtain

$$\widetilde{Q}^k P_\rho^0(t, t_0, \cdot)(s_{kj}) = 0 \quad \text{for } t \in [t_0, T].$$
(3.18)

By virtue of the irreducibility of $\widetilde{Q}^k$, we obtain $P_\rho^0(t, t_0, s_{kj}) = P_\rho^0(t, t_0, k)$, i.e., it is independent of j. Next, we obtain the limit of the Riccati equations. Multiplying both sides of (3.17) by ν_j^k and summing over the index j lead to

$$\sum_{j=1}^{m_k} \nu_j^k P_\rho^\varepsilon(t, t_0, s_{kj}) = \sum_{j=1}^{m_k} \nu_j^k P_\rho^\varepsilon(t_0, t_0, s_{kj})$$
$$- \rho \int_{t_0}^t [\sum_{j=1}^{m_k} \nu_j^k P_\rho^\varepsilon(r, t_0, s_{kj}) G_c(r, s_{kj}) P_\rho^\varepsilon(r, t_0, k)] dr$$
$$+ \rho \int_{t_0}^t [\sum_{j=1}^{m_k} \nu_j^k G_o(r, s_{kj}) + \sum_{j=1}^{m_k} \nu_j^k Q^\varepsilon P_\rho^\varepsilon(r, t_0, \cdot)(s_{kj})] dr.$$

Letting $\varepsilon \to 0$, and using $P_\rho^\varepsilon(t, t_0, s_{kj}) \to P_\rho^0(t, t_0, k)$ uniformly, we obtain

$$\sum_{j=1}^{m_k} \nu_j^k \widehat{Q} \mathbb{1}_{m_k} P_\rho^0(t, t_0, \cdot)(k) = \overline{Q} P_\rho^0(t, t_0, \cdot)(k).$$

Noting $\sum_{j=1}^{m_k} \nu_j^k = 1$, we have

$$P_\rho^0(t, t_0, k) = P_\rho^0(t_0, t_0, k) - \rho \int_{t_0}^t [P_\rho^0(r, t_0, k) \overline{G}_c(r, k) P_\rho^0(r, t_0, k)] dr$$
$$+ \rho \int_{t_0}^t [\overline{G}_o(r, k) + \widehat{Q} P_\rho^0(r, t_0, \cdot)(k)] dr.$$

By virtue of the uniqueness of the solution, we obtain $P_\rho^0(t, t_0, k) = \overline{P}_\rho(t, t_0, k)$. □

As a consequence of the above theorem and the result in Section 2, we obtain the following result.

Theorem 3.7. *Under the conditions of Theorem* 3.2, *for each* $k = 1, \ldots, l$ *and* $j = 1, \ldots, m_j$,

$$\lim_{\varepsilon \to 0} |P_\rho^\varepsilon(t, t_0, s_{kj}) - \overline{P}(t, t_0, k)| \leq |\overline{P}_\rho(t_0, k) - \overline{P}(t_0, k)|e^{-\rho\gamma_{3,0}(t-t_0)} + \frac{\gamma_{5,0}}{\rho\gamma_{3,0}}, \tag{3.19}$$

and the estimate is uniformly for t *in any compact interval.*

Proof. The proof is a combination of the argument used in the last section and Theorem 3.2. We merely note that by the familiar triangle inequality,

$$|P_\rho^\varepsilon(t, t_0, s_{kj}) - \overline{P}(t, k)| \leq |P_\rho^\varepsilon(t, t_0, s_{kj}) - \overline{P}_\rho(t, t_0, k)| + |\overline{P}_\rho(t, t_0, k) - \overline{P}(t, k)|.$$

Taking limit as $\varepsilon \to 0$, the desired result follows. □

4 Numerical Examples

In this section, we present two numerical examples, which demonstrate how our algorithms perform. Suppose we have a two-state Markov chain with generator (3.1), where

$$\widetilde{Q} = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}, \quad \widehat{Q} = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}. \tag{4.1}$$

Set $t_0 = 0$ and $T = 6$.

Consider (2.2) satisfying conditions (A1) and (A2) with the Grammians,

$$\begin{aligned} G_c(t, 1) &= 4 + \sin t, \quad G_c(t, 2) = 3 + \cos(2t); \\ G_o(t, 1) &= 2 + \cos^2 t, \quad G_o(t, 2) = 6 + 2\cos t. \end{aligned} \tag{4.2}$$

Choosing an initial point far away from the balanced realization, Figure 1 shows that the results obtained by solving the algebraic equations and the coupled Riccati differential equations are quite close for large t. We plot $P^\varepsilon(t, 1)$, $P^\varepsilon(t, 2)$, $P_\rho^\varepsilon(t, t_0, 1)$, and $P_\rho^\varepsilon(t, t_0, 2)$, using solid line, dashed line, dotted line, and dash-dot line, respectively. With the increase in ρ, the results improve. Figure 2 illustrates that if the Markov chain changes rapidly, the system of Riccati equations can be replaced by an averaged system, in which the coefficients are averaged out with respect to the stationary measures. We plot $P_\rho^\varepsilon(t, t_0, 1)$, $P_\rho^\varepsilon(t, t_0, 2)$,

and $\overline{P}_\rho(t, t_0, k)$ using dotted line, dash-dotted line, solid line, respectively. Different ε are used; they all show similar results.

Next, consider another system that is nearly uncontrollable at $t = (3\pi)/2$ with the following Grammians,

$$
\begin{aligned}
G_c(t, 1) &= 1.01 + \sin t, \quad G_c(t, 2) = 1.01 + \cos(2t) \\
G_o(t, 1) &= 2 + \cos^2 t, \quad G_o(t, 2) = 6 + 2\cos t.
\end{aligned}
\tag{4.3}
$$

In Figure 3 and Figure 4, we find that our algorithm still works well for this nearly uncontrollable case.

5 Concluding Remarks

This work developed a framework for balanced realizations of linear systems modulated by a continuous-time Markov chain with a finite state space. To reduce computational complexity, we used a ρ-balanced realization. When the state space of the Markov chain is large, we introduced an additional two-time-scale formulation. Decomposition and aggregation methods were used and averaging techniques were utilized. With the balancing developed in this paper, there can be a model reduction by working with diagonally balanced versions of the systems and setting the smallest singular values to zero. This idea is very useful to treat the case when these singular values are not exactly zero but small and close to zero, which is one of the key reasons for the balanced realization.

A Appendix

To prove the desired result, we first state a preliminary result. Then the lemma is proved

A.1 Preliminary Result

Recall that a square matrix $\widehat{B}$ is similar to another square matrix $\widehat{A}$ with the same dimension, if there is a nonsingular transformation Γ such that $\widehat{A} = \Gamma^{-1}\widehat{B}\Gamma$.

Lemma A.1. *For any real square matrix $\widehat{A}$, with real eigenvalues and nonsingular, we have If $\widehat{A}$ is a real symmetric matrix and $|\widehat{A}| \geq b > -\infty$, then $\widehat{A} \geq bI$ or $\widehat{A} \leq -bI$. If $\widehat{A}$ is a real symmetric positive definite matrix, and $0 \leq a \leq |\widehat{A}| \leq b < \infty$, then*

$$
aI \leq \widehat{A} \leq bI.
\tag{A.1}
$$

If $\widehat{A}$ is a real symmetric matrix, positive definite, and $0 < \widehat{A} \leq bI$ for some $b > 0$, then

$$|\widehat{A}| \leq b. \tag{A.2}$$

Proof. The proof is elementary and is omitted. $\quad\square$

We remark that Lemma 3.1 in [9, p. 317] still holds. This together with Lemma A.1 serves as preliminary for the subsequent development.

A.2 Proof of Lemma 2.5

By virtue of (2.11) and the triangle inequality, we obtain

$$|G_c(t)||P(t)|^2 \geq -\sum_{j=1}^{m} |G_j||H_j||P(t)| + |G_o(t)|.$$

Since the quadratic equation

$$|G_c(t)||P(t)|^2 + \sum_{j=1}^{m} |G_j||H_j||P(t)| - |G_o(t)| = 0$$

has one positive and one negative root,

$$|P(t)| \geq \frac{-\sum_{j=1}^{m} |G_j||H_j| + \sqrt{(\sum_{j=1}^{m} |G_j||H_j|)^2 + 4|G_c(t)||G_o(t)|}}{2|G_c(t)|}. \tag{A.3}$$

On the other hand, using (2.11), we have

$$P(t)G_c(t)P(t) = G_o(t) + \sum_{j=1}^{m} G_j P(t) H_j := Z.$$

It is easy to see that

$$|Z| \leq |G_o(t)| + \sum_{j=1}^{m} |G_j||H_j||P(t)|. \tag{A.4}$$

Thus by conditions (A1) and (A2), we have

$$P(t) = G_c^{-1/2}[G_c^{1/2}ZG_c^{1/2}]^{1/2}G_c^{-1/2}, \tag{A.5}$$

and

$$G_c^{1/2}ZG_c^{1/2} \leq G_c|Z| \leq U_c|Z|I.$$

By virtue of [12, Theorem 9, p. 15], $\lambda(AB) = \lambda(BA)$ for nonsingular matrices A and B, we have

$$P(t) \leq G_c^{-1/2}[U_c|Z|I]^{1/2}G_c^{-1/2} = U_c^{1/2}|Z|^{1/2}G_c^{-1} \leq U_c^{1/2}L_c^{-1}|Z|^{1/2}I.$$

Noticing that $U_c^{1/2} L_c^{-1} |Z|^{1/2}$ is a scalar, by virtue of (A.2) and (A.4), we obtain

$$|P(t)| \le U_c^{1/2} L_c^{-1} |Z|^{1/2} \le U_c^{1/2} L_c^{-1} (|G_o(t)| + \sum_{j=1}^{m} |G_j||H_j|)|P(t)|)^{1/2},$$

i.e.,

$$|P(t)|^2 \le U_c L_c^{-2} (|G_o(t)| + \sum_{j=1}^{m} |G_j||H_j|)|P(t)|). \tag{A.6}$$

We also observe that the following equation

$$|P(t)|^2 - U_c L_c^{-2} \sum_{j=1}^{m} |G_j||H_j||P(t)| - U_c L_c^{-2} |G_o(t)| = 0,$$

has one positive and one negative roots. Therefore, solving the corresponding quadratic inequality, we obtain

$$|P(t)| \le \frac{U_c L_c^{-2} \left(\sum_{j=1}^{m} |G_j||H_j| \right) + \sqrt{U_c^2 L_c^{-4} (\sum_{j=1}^{m} |G_j||H_j|)^2 + 4 U_c L_c^{-2} |G_o(t)|}}{2}. \tag{A.7}$$

Combining (A.3) and (A.7),

$$\gamma_2 \le |P(t)| \le \gamma_1, \tag{A.8}$$

where

$$\gamma_1 = \frac{U_c L_c^{-2} \left(\sum_{j=1}^{m} |G_j||H_j| \right) + \sqrt{U_c^2 L_c^{-4} (\sum_{j=1}^{m} |G_j||H_j|)^2 + 4 U_c U_o L_c^{-2}}}{2},$$

$$\gamma_2 = \inf_{x \in [L_c, U_c], y \in [L_o, U_o]} \frac{-\sum_{j=1}^{m} |G_j||H_j| + \sqrt{(\sum_{j=1}^{m} |G_j||H_j|)^2 + 4xy}}{2y}.$$

Since the above function is continuous on compact sets, the positive constant γ_2 is well defined. It follows that

$$\lambda_{\min}(P(t)G_c(t)) \ge \lambda_{\min}(P(t))\lambda_{\min}(G_c(t)) = \gamma_2 L_c := \gamma_3,$$

$$\lambda_{\max}(P(t)G_c(t)) \le \lambda_{\max}(P(t))\lambda_{\max}(G_c(t)) = \gamma_1 U_c := \gamma_4.$$

Next, we want to estimate $\dot{P} = \dot{P}(t)$. By differentiating (2.11), we have

$$\dot{P}G_c P + P G_c \dot{P} + (P \dot{G}_c P - \dot{G}_o) - \sum_{j=1}^{m} G_j \dot{P} H_j = 0, \tag{A.9}$$

i.e.,

$$\dot{P}G_c P + P G_c \dot{P} + (P \dot{G}_c P - \dot{G}_o) = \sum_{j=1}^{m} G_j \dot{P} H_j. \tag{A.10}$$

21

Using the Vec operator and Kronecker product [7] and noting $Vec(AXB) = (B' \otimes A)VecX$ and $Vec(AX + XB) = (I_m \otimes A + B' \otimes I_n)VecX$, we have

$$(I_m \otimes (PG_c) + I_m \otimes (G_cP))Vec\dot{P} + Vec(P\dot{G}_cP - \dot{G}_o) = \sum_{j=1}^{m}(H_j' \otimes G_j)Vec\dot{P}. \quad (A.11)$$

We obtain

$$(I_m \otimes (PG_c) + I_m \otimes (G_cP) - \sum_{j=1}^{m}(H_j' \otimes G_j))Vec\dot{P} = Vec(\dot{G}_o - P\dot{G}_cP).$$

It follows that

$$|\dot{P}|_2 = |Vec\dot{P}|_2 \leq |Vec(\dot{G}_o - P\dot{G}_cP)|_2|(I_m \otimes (PG_c) + I_m \otimes (G_cP) - \sum_{j=1}^{m}(H_j' \otimes G_j))^{-1}|_2, \quad (A.12)$$

where $|\cdot|_2$ denote the Euclidean norm ($|v|_2 = \sqrt{\sum_{i=1}^{r}\sum_{j=1}^{l}v_{ij}^2}$ for any $v \in \mathbb{R}^{r \times l}$). Thus

$$|\dot{P}| \leq |(\dot{G}_o - P\dot{G}_cP)||(I_m \otimes (PG_c) + I_m \otimes (G_cP) - \sum_{j=1}^{m}(H_j' \otimes G_j))^{-1}| := \gamma_5. \quad (A.13)$$

Using a similar argument as in [9], we can show that $|(\dot{G}_o - P\dot{G}_cP)|$ is bounded. Write

$$X = (I_m \otimes (PG_c) + I_m \otimes (G_cP) - \sum_{j=1}^{m}(H_j' \otimes G_j)).$$

By virtue of Cayley-Hamilton theorem, we obtain

$$X^{-1} = c_0I + c_1X + \cdots + c_{n-1}X^{n-1}, \quad (A.14)$$

where c_i are some constants. Note the eigenvalues of $A \otimes B$ are $\lambda_i(A)\lambda_j(B)$, where $\lambda_i(A)$ and $\lambda_j(B)$ denote the eigenvalues of A and B, respectively. Thus, we obtain

$$|X| \leq 2|PG_c| + \sum_{j=1}^{m}|(H_j' \otimes G_j)| < \infty.$$

By (A.14), using the triangle inequality together with Lemma A.1, we have

$$|\dot{P}| < \gamma_5.$$

By virtue of (A.1), we have

$$-\gamma_5I \leq \dot{P}(t) \leq \gamma_5I.$$

The proof is completed.   □

References

[1] H. Abou-Kandil, G. Freiling, V. Ioescu, and G. Jank *Matrix Riccati Equations in Control and Systems Theory*, Birkhäuser, Basel, Boston, Berlin, 2003.

[2] B.D.O. Anderson and J.B. Moore *Optimal Control: Linear Quadratic Methods*, Prentice Hall, Englewood Cliffs, NJ, 1990.

[3] D.P. Bersekas, *Dynamic Programming and Stochastic Control*, Academic Press, New York, 1976.

[4] P.J. Courtois, *Decomposability: Queueing and Computer System Applications*, Academic Press, New York, NY, 1977.

[5] W.H. Fleming and R.W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, New York, NY, 1975.

[6] K. Glover, All optimal Hankel norm approximations of linear multivariable systems and their L-error bounds, *Internat. J. Control*, **39** (1984), 1145-1193.

[7] A. Graham, *Kronecker Product and Matrix Calculus with Applications*, Ellis Horwood Ltd., Chechester, 1981.

[8] U. Helmke and J.B. Moore, *Optimization and Dynamical systems*, Springer-Verlag, New York, NY, 1994.

[9] J. Imae, J.E. Perkins, and J.B. Moore, Toward time-varying balanced realization via Riccati equations, *Math. Control Signals Systems*, **5** (1992), 313–326.

[10] Y. Ji and H.J. Chizeck, Jump linear quadratic Gaussian control in continuous-time, *IEEE Trans. Automat. Control*, **37** (1992), 1884–1892.

[11] M. Mariton and P. Bertrand, Robust jump linear quadratic control: A mode stabilizing solution, *IEEE Trans. Automat. Control*, **30** (1985), 1145–1147.

[12] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, J. Wiley, Chichester, England, 1999.

[13] B.C. Moore, Principal component in linear systems: Controllability, observability, and model reduction, *IEEE Trans. Automatic Control*, **26** (1981), 17-31.

[14] J.E. Perkins, U. Helmke, and J.B. Moore, Balanced realizations via gradient flow techniques, *Sys. Control Lett.*, **14** (1990), 369–379.

[15] H. Sandberg and A. Rantzer, Balanced truncation of linear time-varying systems, *IEEE Trans. Automat. Control*, **49** (2004), 217–229.

[16] S. Shokoohi, L.M. Silverman, and P.M. Van Dooren, Linear time variable systems: Balancing and model reduction, *IEEE Trans. Automat. Control*, **28** (1983), 810-822.

[17] S.P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser, Boston, MA, 1994.

[18] H.A. Simon and A. Ando, Aggregation of variables in dynamic systems, *Econometrica*, **29** (1961), 111–138.

[19] E.I. Verriest and T. Kailath, On generalized balanced realizations, *IEEE Trans. Automat. Control*, **28** (1983), 833-844.

[20] G. Yin and S. Dey, Weak convergence of hybrid filtering problems involving nearly completely decomposable hidden Markov chains, *SIAM J. Control Optim.*, **41** (2003), 1820-1842.

[21] G. Yin, V. Krishnamurthy, and C. Ion, Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization, *SIAM J. Optim.*, **14** (2004), 1187-1215.

[22] G. Yin, Y.J. Liu, and H. Yang, Bounds of ruin probability for regime-switching models using time scale separation, to appear in *Scandinavian Actuarial Journal*.

[23] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*, Springer-Verlag, New York, 1998.

[24] G. Yin and Q. Zhang, *Discrete-time Markov Chains: Two-time-scale Methods and Applications*, Springer, New York. 2005.

[25] G. Yin, Q. Zhang, and G. Badowski, Asymptotic properties of a singularly perturbed Markov chain with inclusion of transient states, *Ann. Appl. Probab.* **10** (2000), 549-572.

[26] J. Yong and X.Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer-Verlag, New York, 1999.

[27] Q. Zhang and G. Yin, On nearly optimal controls of hybrid LQG problems, *IEEE Trans. Automat. Control*, **44** (1999) 2271–2282.

[28] X.Y. Zhou and G. Yin, Markowitz's mean-variance portfolio selection with regime switching: A Continuous-time model, *SIAM J. Control Optim*, **42** (2003), 1466–1482.
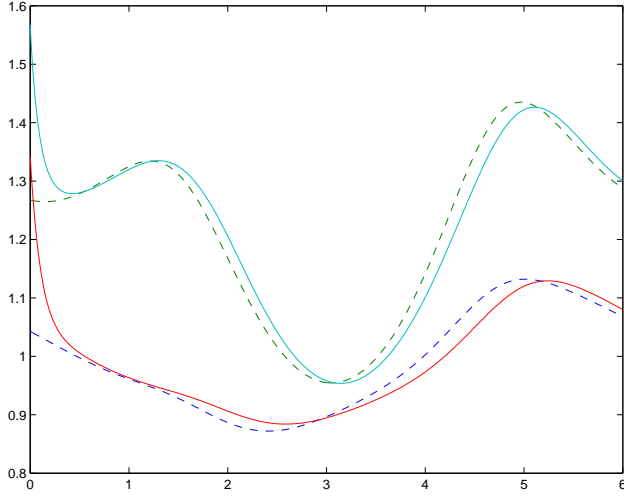
(a) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.1, \rho = 1$
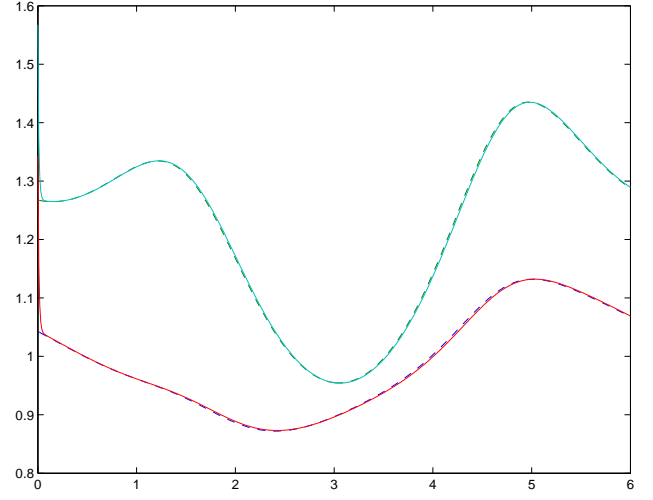
(b) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.1, \rho = 10$
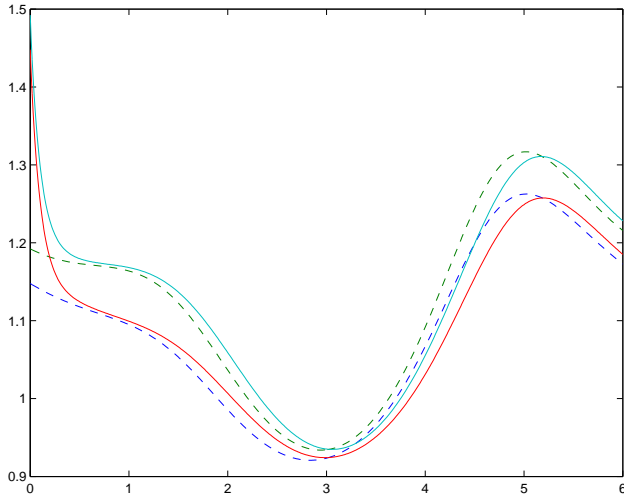
(c) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.01, \rho = 1$
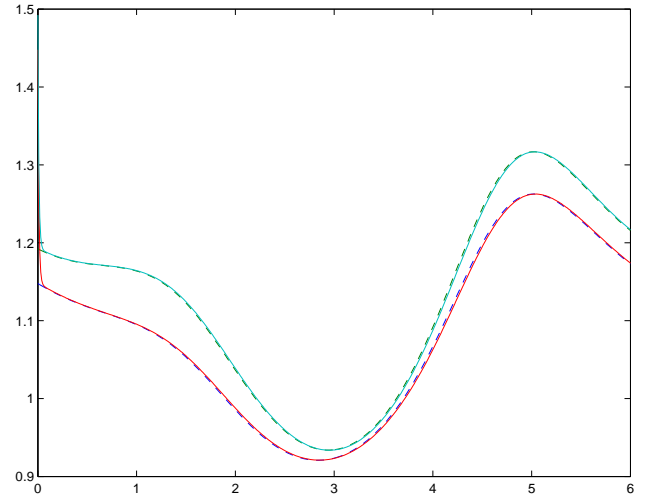
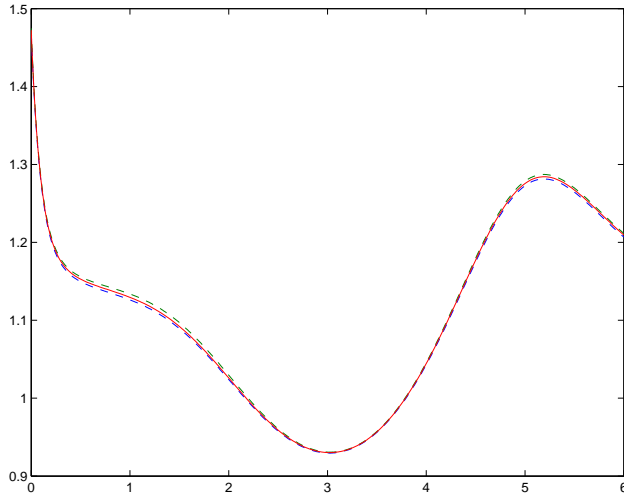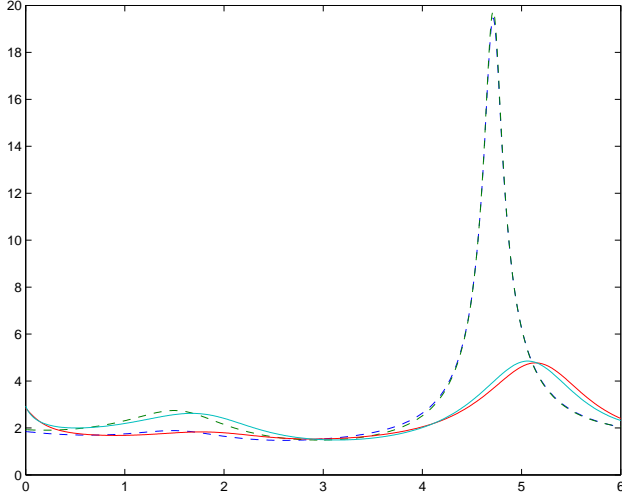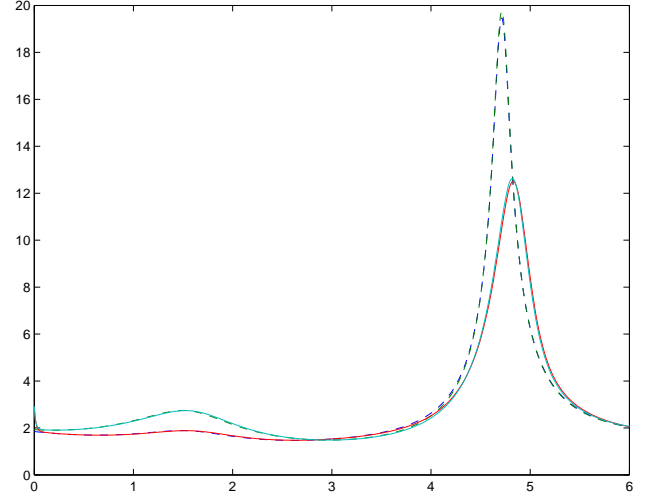(d) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.01, \rho = 10$

Figure 1: $\{P^\varepsilon(t,i)\}$ is the solution of system of coupled algebraic equations and $\{P^\varepsilon_\rho(t,t_0,i)\}$ is the solution of the system of coupled Riccati differential equations.
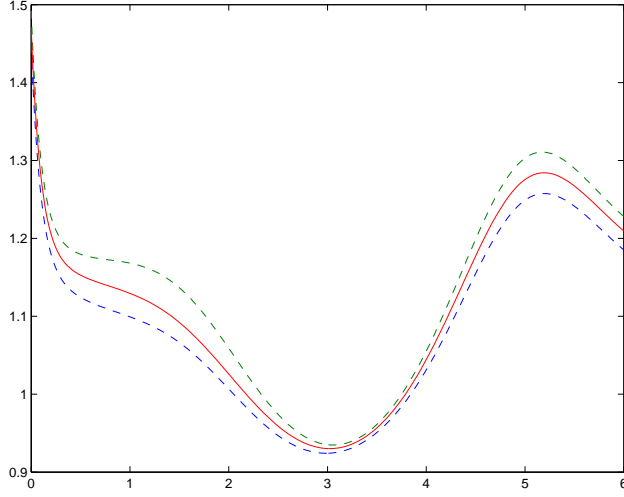
(a) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.01, \rho = 1$
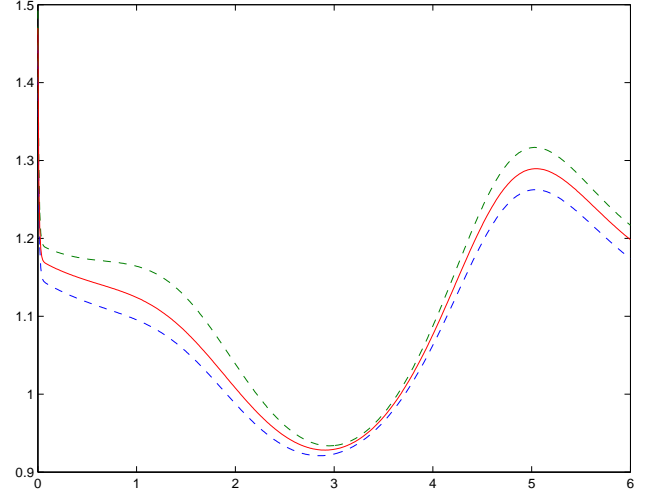
(b) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.01, \rho = 10$
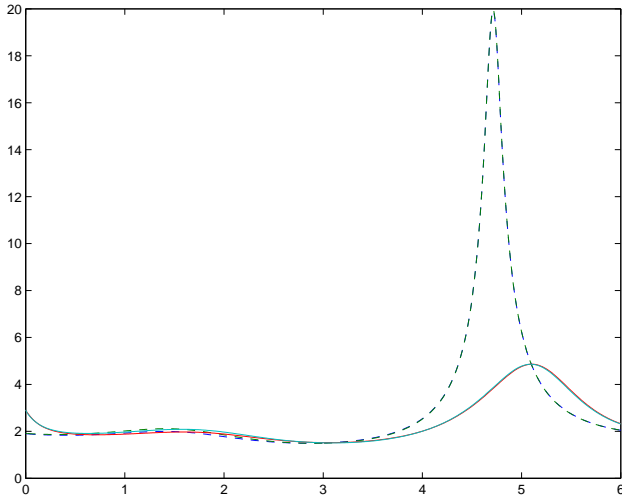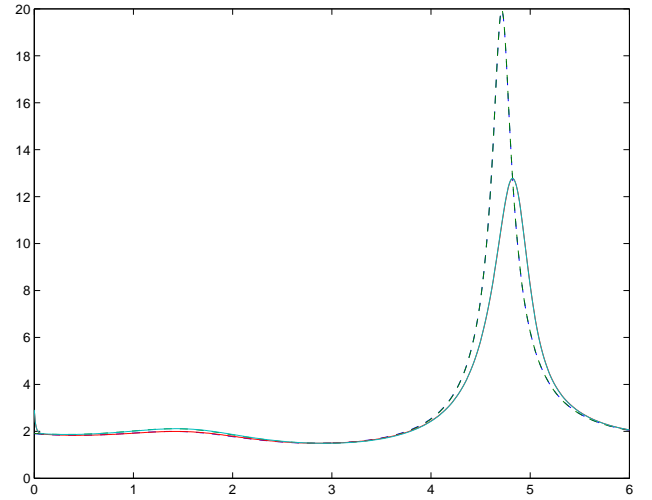
(c) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.001, \rho = 1$
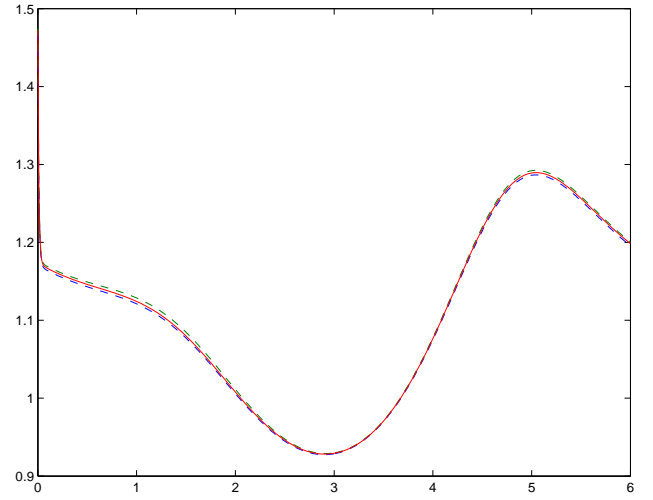
(d) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.001, \rho = 10$

Figure 2: $\{P_\rho^\varepsilon(t, t_0, i)\}$ is the solution of the system of coupled Riccati differential equations and $\overline{P}_\rho(t, t_0, k)$ is the solution of the Riccati differential equation whose coefficients are averaged out with respect to the stationary measures.

(a) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.1, \rho = 1$

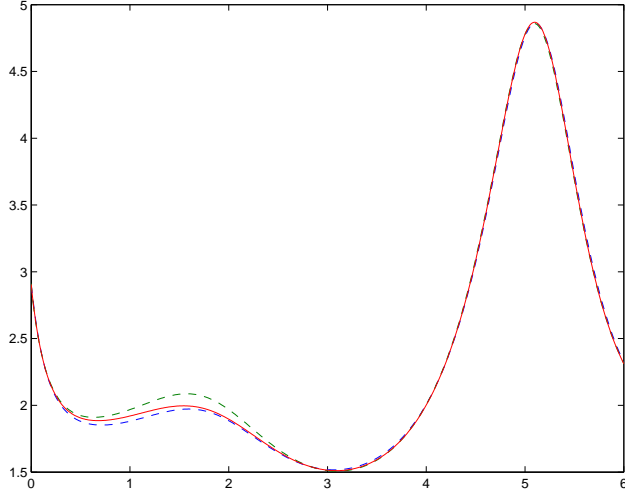(b) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.1, \rho = 10$

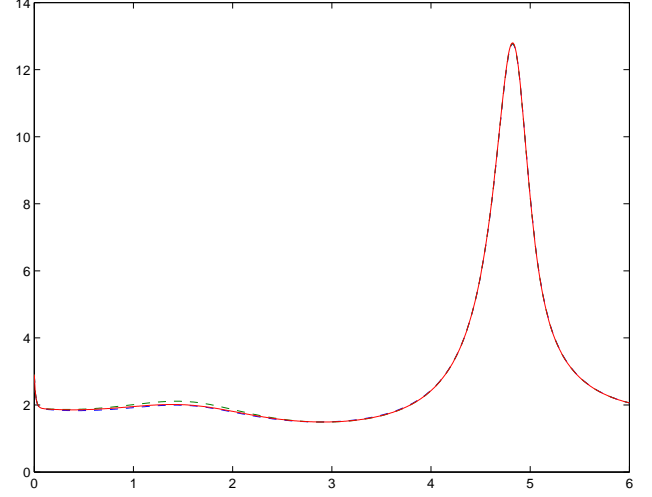(c) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.01, \rho = 1$

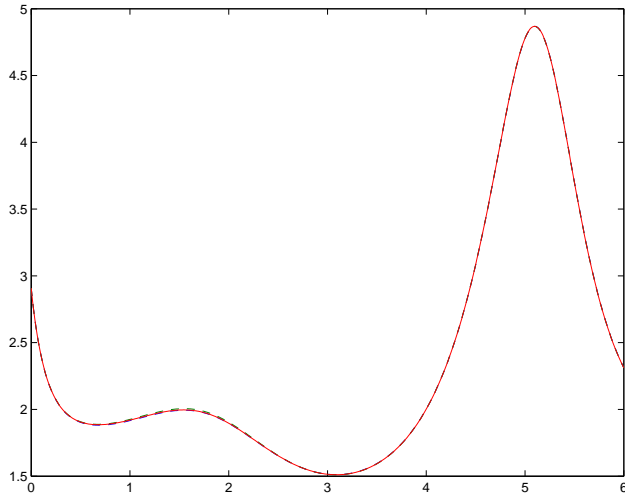(d) $P^\varepsilon(t,i)$ vs. $P^\varepsilon_\rho(t,t_0,i)$ with $\varepsilon = 0.01, \rho = 10$

Figure 3: For the nearly uncontrollable system, $\{P^\varepsilon(t,i)\}$ is the solution of the system of coupled algebraic equations and $\{P^\varepsilon_\rho(t,t_0,i)\}$ is the solution of the system of the coupled Riccati differential equations.
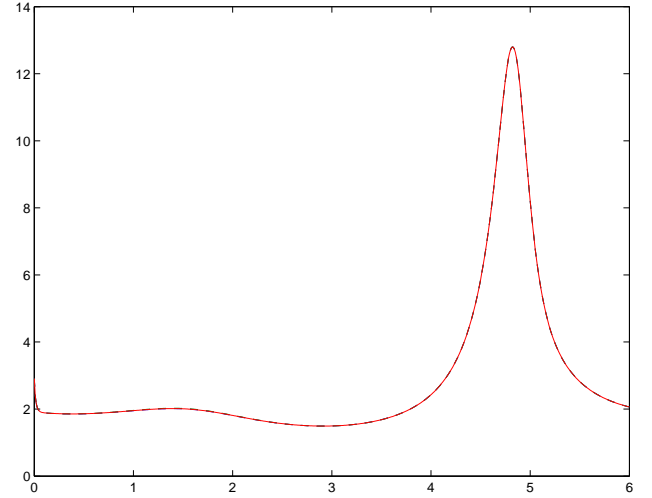
(a) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.01, \rho = 1$

(b) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.01, \rho = 10$

(c) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.001, \rho = 1$

(d) $P_\rho^\varepsilon(t, t_0, i)$ vs. $\overline{P}_\rho(t, t_0, k)$ with $\varepsilon = 0.001, \rho = 10$

Figure 4: For nearly uncontrollable system, $\{P_\rho^\varepsilon(t, t_0, i)\}$ is the solution of the system of coupled Riccati differential equations and $\overline{P}_\rho(t, t_0, k)$ is the solution of the Riccati differential equation whose coefficients are averaged out with respect to the stationary measures.