# Efficient Suboptimal Solutions of Switched LQR Problems

Wei Zhang, Alessandro Abate and Jianghai Hu

*Abstract*— **This paper studies the discrete-time switched LQR (DSLQR) problem using a dynamic programming approach. Based on some nice properties of the value functions, efficient algorithms are proposed to solve the finite-horizon and infinite-horizon suboptimal DSLQR problems. More importantly, we establish analytical conditions under which the strategies generated by the algorithms are stabilizing and suboptimal. These conditions are derived explicitly in terms of subsystem matrices and are thus very easy to verify. The proposed algorithms and the analysis provide a systematical way of solving the DSLQR problem with guaranteed close-loop stability and suboptimal performance. Simulation results indicate that the proposed algorithms can efficiently solve not only specific but also randomly generated DSLQR problems, making NP-hard problems numerically tractable.**

## I. Introduction

Optimal control of switched systems has many practical applications [1], [2] and has challenged researchers for many years. The bottleneck mostly lies in the determination of the optimal switching strategy. Many methods have been proposed to tackle this problem, most of which are in a divide-and-conquer manner. Algorithms for optimizing the switching instants for a fixed mode sequence have been developed for general switched systems in [3] and for switched systems with autonomous dynamics in [4]. Although an algorithm for updating the switching sequence is discussed in [4], finding the best switching sequence is still an NP-hard problem, even for switched linear systems.

This paper studies the discrete-time quadratic regulation problem for switched linear systems (DSLQR) based on a dynamic programming (DP) approach. A generic way of solving a DP problem is by gridding the state space. Such method has been used to study various optimal control problems of switched systems [5], [6], [7]. Its main drawback lies in the exponential growth of the complexity as the state dimension increases. Fortunately, for some simple classes of problems, the value functions may have some nice analytical properties that can be used to simplify the numerical computation. For this reason, the quadratic optimal control problem of switched linear/affine systems has attracted many research attentions [8], [9], [10]. In our previous papers [11], [12], we showed that the value function of the DSLQR problem is the pointwise minimum over a finite number of quadratic functions. More importantly, we showed that these quadratic functions are exactly characterized by a set of positive semidefinite matrices which can be obtained using the difference Riccati equation. These properties can dramatically simplify the solution of the DSLQR problem.

The main contribution of this paper is the design and analysis of various efficient algorithms for solving the finite-horizon and infinite-horizon suboptimal DSLQR problems. The key idea is to use convex optimization to identify and remove the matrices that are redundant in terms of characterizing the optimal and suboptimal strategies. This is in line with the approaches of Neuro-dynamic programming ([13]) and approximate dynamic programming ([14]), which both try to simplify the computation by finding a compact representation of the value functions up to certain numerical relaxations. Compared with the previous work, our distinctions are the following. (i) We introduce the new concepts of algebraic redundancy and numerical redundancy to set up a formal framework for analyzing the suboptimal algorithms. (ii) A rigorous analysis is carried out on the propagation of the approximation errors through the value iterations. (iii) More importantly, we establish conditions under which the strategies generated by the proposed algorithms are stabilizing and suboptimal. Furthermore, these conditions are derived explicitly in terms of subsystem matrices and are very easy to verify. Therefore, the proposed algorithms and the analysis provide a systematical way of solving the DSLQR problem with guaranteed close-loop stability and suboptimal performance. Simulation results indicate that the proposed algorithms can efficiently solve not only specific but also randomly generated DSLQR problems, making NP-hard problems numerically tractable.

## II. Problem Formulation

Consider the discrete-time switched linear system described by:

$$x(t+1)=A_{v(t)}x(t)+B_{v(t)}u(t), t \in T_N \triangleq \{0,\dots,N-1\}. \quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the continuous state, $v(t) \in \mathbb{M} \triangleq \{1,\dots,M\}$ is the discrete mode, $u(t) \in \mathbb{R}^p$ is the continuous control and $T_N$ is the control horizon with length $N$ (possibly infinite). The sequence of pairs $\{(u(t),v(t))\}_{t=0}^{N-1}$ is called the *hybrid control sequence*. For each $i \in \mathbb{M}$, $A_i$ and $B_i$ are constant matrices of appropriate dimensions and the pair $(A_i,B_i)$ is called a subsystem of (1). This switched linear system is time invariant in the sense that the set of available subsystems $\{(A_i,B_i)\}_{i=1}^{M}$ is independent of time $t$. We assume that there is no internal forced switchings, i.e., the

This work was partially supported by the National Science Foundation under Grant CNS-0643805

W. Zhang and J. Hu are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN47906. Email:{zhang70,jianghai@purdue.edu}.

A. Abate is with the Department of Aeronautics and Astronautics, Stanford University, CA 94305. Email:{aabate@stanford.edu}.

system can stay at or switch to any mode at any time instant. At each time $t \in T_N$, denote by $\xi_{t,N} \triangleq (\mu_{t,N}, \nu_{t,N}) : \mathbb{R}^n \to \mathbb{R}^p \times \mathbb{M}$ the *hybrid control law* of system (1), where $\mu_{t,N} : \mathbb{R}^n \to \mathbb{R}^p$ is called the *continuous control law* and $\nu_{t,N} : \mathbb{R}^n \to \mathbb{M}$ is called the *switching control law*. A sequence of hybrid control laws over the horizon $T_N$ constitutes an $N$-*horizon feedback policy*: $\pi_N \triangleq \{\xi_{0,N}, \xi_{1,N}, \ldots, \xi_{N-1,N}\}$. If system (1) is driven by a feedback policy $\pi_N$, then the closed-loop dynamics is governed by

$$x(t+1) = A_{\nu_{t,N}(x(t))} x(t) + B_{\nu_{t,N}(x(t))} \mu_{t,N}(x(t)), t \in T_N. \quad (2)$$

For a given initial state $x(0) = z$, the performance of the feedback policy $\pi_N$ can be measured by the following cost functional:

$$J_{\pi_N}(z) = \psi(x(N)) + \sum_{t=0}^{N-1} L(x(t), \mu_{t,N}(x(t)), \nu_{t,N}(x(t))), \quad (3)$$

where $\psi : \mathbb{R}^n \to \mathbb{R}^+$ and $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{M} \to \mathbb{R}^+$ are called the *terminal cost function* and the *running cost function*, respectively. In this paper, the functions $\psi$ and $L$ are assumed to take the following quadratic forms:

$$\psi(x) = x^T Q_f x, \quad L(x, u, v) = x^T Q_v x + u^T R_v u,$$

for $x \in \mathbb{R}^n$, $u \in \mathbb{R}^p$, $v \in \mathbb{M}$, where $Q_f = Q_f^T \succeq 0$ is the terminal state-weighting matrix, and $Q_v = Q_v^T \succeq 0$ and $R_v = R_v^T \succ 0$ are the running weighting matrices for the state and the control, respectively, for subsystem $v \in \mathbb{M}$. When the control horizon $N$ is infinite, the terminal cost will never be incurred and the objective function becomes:

$$J_{\pi_\infty}(z) = \sum_{t=0}^{\infty} L(x(t), \mu_{t,\infty}(x(t)), \nu_{t,\infty}(x(t))). \quad (4)$$

For a possibly infinite positive integer $N$, denote by $\Pi_N$ the set of all admissible $N$-horizon policies, i.e., the set of all sequence of functions $\pi_N = \{\xi_{0,N}, \ldots, \xi_{N-1,N}\}$ with $\xi_{t,N} : \mathbb{R}^n \to \mathbb{R}^p \times \mathbb{M}$ for $t \in T_N$. The goal of this paper is to find the optimal policy $\pi_N^*$ that minimizes the quadratic cost function defined in (3) or (4). This problem is a natural extension of the classical LQR problem to the switched linear system case and is thus called the *Discrete-time Switched LQR problem*, hereby referred to as the DSLQR problem.

*Problem 1 (DSLQR problem):* For a given initial state $z \in \mathbb{R}^n$ and a possibly infinite positive integer $N$, find the $N$-horizon policy $\pi_N \in \Pi_N$ that minimizes $J_{\pi_N}(z)$ subject to the dynamic equation (2).

To solve Problem 1, for each time $t \in T_N$, we define the value function $V_{t,N} : \mathbb{R}^n \to \mathbb{R}$ as:

$$V_{t,N}(z) = \inf_{\substack{u(j) \in \mathbb{R}^p, v(j) \in \mathbb{M} \\ t \le j \le N-1}} \Big\{ \psi(x(N)) + \sum_{j=t}^{N-1} L(x(j), u(j), v(j)) \Big| \text{ subject to eq. (1) with } x(t) = z \Big\}. \quad (5)$$

The $V_{t,N}(z)$ so defined is the minimum cost-to-go starting from state $z$ at time $t$. The minimum cost for the DSLQR problem with an initial condition $x(0) = x_0$ is simply $V_{0,N}(x_0)$. Due to the time-invariant nature of the switched system (1), its value function depends only on the number of remaining time steps, i.e.,

$$V_{t,N}(z) = V_{t+m,N+m}(z),$$

for all $z \in \mathbb{R}^n$ and all integers $m \ge -t$. In the rest of this paper, when no ambiguity arises, we will denote by $V_k(z) \triangleq V_{N-k,N}(z)$ and $\xi_k \triangleq \xi_{N-k,N}$ the value function and the hybrid control law, respectively, at time $t = N - k$ when there are $k$ time steps left. With the new notations, the $N$-horizon policy $\pi_N$ can also be written as $\pi_N = \{\xi_N, \ldots, \xi_1\}$. For any positive integer $k$, the newly introduced $\xi_k$ can be thought of as the first step of a $k$-horizon policy.

By a standard result of Dynamic Programming [15], for any finite integer $N$, the value function $V_N$ can be obtained recursively using the one-stage *value iteration*:

$$V_{k+1}(z) = \inf_{u,v} \{L(z, u, v) + V_k(A_v z + B_v u)\}, \forall z \in \mathbb{R}^n,$$

with the initial condition $V_0(z) = \psi(z)$, $\forall z \in \mathbb{R}^n$. Denote by $V_\infty(\cdot)$ the pointwise limit (if it exists) of the sequence of functions $\{V_k(\cdot)\}_{k=0}^\infty$ generated by the value iterations. It is well known ([15]) that even if $V_\infty(z)$ exists, it may not always coincide with the true infinite-horizon value function. To emphasize its substantial difference from the finite-horizon value function, the infinite-horizon value function is specially denoted by $V^*(z)$, i.e., $V^*(z) = \inf_{\pi_\infty \in \Pi_\infty} J_{\pi_\infty}(z)$.

## III. The Value Function

In this section, we review some important properties of the value functions of the DSLQR problems, which have been derived in our previous papers [16], [12]. These properties are crucial for the design and analysis of the suboptimal algorithms to be developed in Sections IV and V.

In the special case when $M = 1$, the DLQRS problem degenerates into the classical LQR problem. In this degenerate case, we denote by $(A, B)$ the subsystem and by $Q$ and $R$ the state and control weighting matrices. Then, according to the LQR theory, the value function defined in (5) is of the following quadratic form:

$$V_k(z) = z^T P_k z, \quad k = 0, \ldots, N, \quad (6)$$

where $\{P_k\}_{k=0}^N$ is a sequence of positive semi-definite (p.s.d.) matrices satisfying the Difference Riccati Equation (DRE)

$$P_{k+1} = Q + A^T P_k A - A^T P_k B (R + B^T P_k B)^{-1} B^T P_k A, \quad (7)$$

with initial condition $P_0 = Q_f$.

In general, when $M \ge 2$, the value function is no longer of a simple quadratic form as in (6). Nevertheless, the DRE can be generalized to the Switched LQR problems. Let $\mathcal{A}$ be the set of all p.s.d. matrices. The DRE (7) can be viewed as a mapping from $\mathcal{A}$ to $\mathcal{A}$ depending on the matrices $(A, B, Q, R)$. We call this mapping the *Riccati Mapping* and

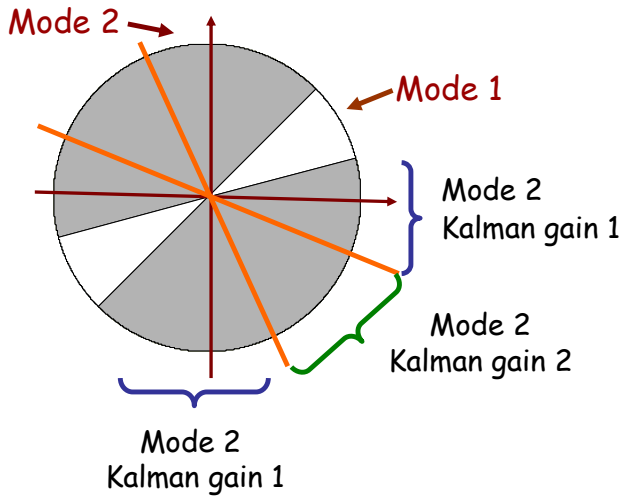


Fig. 1. Typical optimal decision regions of a 2-switched system, where mode 1 is optimal within the white region and mode 2 is optimal within the gray region. The optimal mode region is further divided into smaller conic regions, each of which corresponds to a different Kalman gain.

denote by $\rho_i : \mathcal{A} \to \mathcal{A}$ the Riccati Mapping of subsystem $i \in \mathbb{M}$, i.e.,

$$\rho_i(P)=Q_i+A_i^T PA_i-A_i^T PB_i(R_i+B_i^T PB_i)^{-1}B_i^T PA_i. \quad (8)$$

*Definition 1:* Let $2^{\mathcal{A}}$ be the power set of $\mathcal{A}$. The mapping $\rho_{\mathbb{M}} : 2^{\mathcal{A}} \to 2^{\mathcal{A}}$ defined by:

$$\rho_{\mathbb{M}}(\mathcal{H}) = \{\rho_i(P) : \text{for some } i \in \mathbb{M} \text{ and } P \in \mathcal{H}\}$$

is called the *Switched Riccati Mapping* (SRM) of Problem 1.

In words, the SRM maps a **set** of p.s.d. matrices to another **set** of p.s.d. matrices and each matrix in $\rho_{\mathbb{M}}(\mathcal{H})$ is obtained by taking the classical Riccati mapping of some matrix in $\mathcal{H}$ through some subsystem $i \in \mathbb{M}$.

*Definition 2:* The sequence of sets $\{\mathcal{H}_k\}_{k=0}^N$ generated iteratively by $\mathcal{H}_{k+1} = \rho_{\mathbb{M}}(\mathcal{H}_k)$ with initial condition $\mathcal{H}_0 = \{Q_f\}$ is called the *Switched Riccati Sets* (SRSs) of Problem 1.

An important fact about the DSLQR problem is that its value functions are completely characterized by the SRSs.

*Theorem 1 ([11]):* The value function for the DSLQR problem at time $N-k$, i.e., with $k$ time steps left, is

$$V_k(z) = \min_{P\in\mathcal{H}_k} z^T Pz. \quad (9)$$

Furthermore, for $k = 1, \ldots, N$, if we define

$$(P_k^*(z), i_k^*(z)) = \underset{(P\in\mathcal{H}_{k-1}, i\in\mathbb{M})}{\arg\min} \; z^T \rho_i(P) z, \quad (10)$$

then the optimal hybrid control law at time $t = N-k$ is $\xi_k^*(z) = (\mu_k^*(z), \nu_k^*(z))$, where $\mu_k^*(z) = -K_{i_k^*(z)}(P_k^*(z))z$ and $\nu_k^*(z) = i_k^*(z)$. Here, $K_i(P)$ is the Kalman gain for subsystem $i$ with matrix $P$, i.e.,

$$K_i(P) \triangleq (R_i + B_i^T PB_i)^{-1} B_i^T PA_i. \quad (11)$$

.

Compared with the discrete-time LQR case, the value function of the DSLQR problem is no longer a single quadratic function; it actually becomes the pointwise minimum of a finite number of quadratic functions. In addition, at each time step, instead of having a single Kalman gain for the entire state space, the optimal state feedback gain becomes state dependent. Furthermore, the minimizer $(P_k^*(z), i_k^*(z))$ of equation (10) is radially invariant, indicating that at each time step all the points along the same radial direction have the same optimal hybrid control law. These interesting properties are illustrated in Fig. 1 using an example in $\mathbb{R}^2$ with 2 subsystems. At each time instant, the state space is decomposed into two homogeneous regions: the white region and the gray region, which are called the *optimal switching regions*. Within the white region, one mode, say mode 1, is optimal; within the gray region, the other mode, namely mode 2, is optimal. Furthermore, the states within the same optimal switching region may have different optimal feedback gains (Kalman gains). This is illustrated in Fig. 1 by further dividing the gray regions into smaller conic regions, each of which correspond to a different Kalman gain. It is worth mentioning that in higher dimensional state space, the decision regions are still cones, however, these cones may no longer be convex and there might be complex manifolds defining the boundaries between adjacent cones. A salient feature of the DSLQR problem is that all these complex decision regions are completely characterized by a finite number of matrices in the switched Riccati sets $\{\mathcal{H}_k\}_{k=0}^N$ that can be obtained analytically.

## IV. Suboptimal Control in Finite Horizon

It is clear from Theorem 1 that the key for solving the DSLQR problem is the computation of the SRSs $\{\mathcal{H}_k\}_{k=0}^N$. The main challenge lies in the exponential growth of $|\mathcal{H}_k|$ as $k$ increases. Two important facts about the value functions can be used to simplify the computation of $\{\mathcal{H}_k\}_{k=0}^N$. (i) The matrices in $\mathcal{H}_k$ that make no contribution to the minimum of (9) can be directly removed without causing any error. (ii) When suboptimal performance is acceptable, the matrices in $\mathcal{H}_k$ that make only a "small" contribution to the minimum (9) can also be removed. The goal of this subsection is to use these two ideas to simplify the computation of the (sub)-optimal solution and to analyze the impacts of these simplifications on various aspects of the close-loop system.

### A. Redundancy and Equivalent Subsets

To formalize the above idea, we introduce a few definitions.

*Definition 3 (Algebraic Redundancy):* A matrix $\hat{P} \in \mathcal{H}$ is called *(algebraic) redundant* if for any $z \in \mathbb{R}^n$, there exist a matrix $P \in \mathcal{H}$ such that $z^T Pz \le z^T \hat{P} z$.

If $\hat{P} \in \mathcal{H}$ is redundant, then $\mathcal{H}$ and $\mathcal{H} \setminus \{\hat{P}\}$ will define the same value functions. In this sense, these two sets are equivalent.

*Definition 4 (Equivalent Subset (ES)):* Let $\mathcal{H}$ and $\hat{\mathcal{H}}$ be two sets of p.s.d matrices.

**Algorithm 1**

1) Denote by $P^{(i)}$ the $i^{th}$ matrix in $\mathcal{H}_k$. Specify a tolerance $\epsilon$ and set $\mathcal{H}_k^{(1)} = \{P^{(1)}\}$.
2) For each $i = 2, \ldots, |\mathcal{H}_k|$, if $P^{(i)}$ satisfies the condition in Lemma 1 with respect to $\mathcal{H}_k$, then $\mathcal{H}_k^{(i)} = \mathcal{H}_k^{(i-1)}$; otherwise $\mathcal{H}_k^{(i)} = \mathcal{H}_k^{(i-1)} \cup \{P^{(i)}\}$.
3) Return $\mathcal{H}_k^{(|\mathcal{H}_k|)}$.

1) The set $\hat{\mathcal{H}}$ is called *equivalent* to $\mathcal{H}$, denoted by $\mathcal{H} \sim \hat{\mathcal{H}}$, if $\min_{P\in\mathcal{H}} z^T P z = \min_{\hat{P}\in\hat{\mathcal{H}}} z^T \hat{P} z$, $\forall z \in \mathbb{R}^n$.
2) $\hat{\mathcal{H}}$ is called an *equivalent subset* of $\mathcal{H}$ if $\hat{\mathcal{H}} \subseteq \mathcal{H}$ and $\hat{\mathcal{H}} \sim \mathcal{H}$.

Some matrices in $\mathcal{H}_k$ are "almost" algebraic redundant in the sense that removing them will only introduce a small error to the value function. We call these matrices numerically redundant.

*Definition 5 (Numerical Redundancy):* A matrix $\hat{P} \in \mathcal{H}_k$ is called *(numerically) $\epsilon$-redundant* with respect to $\mathcal{H}_k$ if

$$\min_{P\in\mathcal{H}_k\setminus\hat{P}} z^T P z \le \min_{P\in\mathcal{H}_k} z^T (P+\epsilon I_n) z, \text{ for any } z \in \mathbb{R}^n.$$

*Definition 6 ($\epsilon$-ES):* The set $\mathcal{H}_k^\epsilon$ is called an $\epsilon$-Equivalent-Subset ($\epsilon$-ES) of $\mathcal{H}_k$ if $\mathcal{H}_k^\epsilon \subset \mathcal{H}_k$ and for all $z \in \mathbb{R}^n$,

$$\min_{P\in\mathcal{H}_k} z^T P z \le \min_{P\in\mathcal{H}_k^\epsilon} z^T P z \le \min_{P\in\mathcal{H}_k} z^T (P+\epsilon I_n) z.$$

Removing the $\epsilon$-redundant matrices may introduce some error for the value function; but the error is no larger than $\epsilon$ for $\|z\| \le 1$. To simplify the computation, for a given tolerance $\epsilon$, we want to prune out as many $\epsilon$-redundant matrices as possible. The following lemma provides a sufficient condition for testing the $\epsilon$-redundancy for a given matrix.

*Lemma 1:* $\hat{P}$ is $\epsilon$-redundant in $\mathcal{H}_k$ if there exist nonnegative constants $\alpha_1, \ldots, \alpha_{|\mathcal{H}_k|-1}$ such that $\sum_{i=1}^{|\mathcal{H}_k|-1} \alpha_i = 1$ and $\hat{P} + \epsilon I_n \succeq \sum_{i=1}^{|\mathcal{H}_k|-1} \alpha_i P^{(i)}$, where $\{P^{(i)}\}_{i=1}^{|\mathcal{H}_k|-1}$ is an enumeration of $\mathcal{H}_k \setminus \{\hat{P}\}$.

The condition in Lemma 1 can be easily verified using various existing convex optimization algorithms [17]. To compute an $\epsilon$-ES of $\mathcal{H}_k$, we only need to remove the matrices in $\mathcal{H}_k$ that satisfy the condition in Lemma 1. The detailed procedure is summarized in Algorithm 1. Denote by $Algo_\epsilon(\mathcal{H}_k)$ the $\epsilon$-ES of $\mathcal{H}_k$ returned by the algorithm. The next step is to combine this algorithm with the SRM. To this end, we define $\{\mathcal{H}_k^\epsilon\}_{k=0}^N$ iteratively as:

$$\mathcal{H}_0^\epsilon = \mathcal{H}_0 \text{ and } \mathcal{H}_{k+1}^\epsilon = Algo_\epsilon(\rho_{\mathbb{M}}(\mathcal{H}_k^\epsilon)), \text{ for } k \in T_N. \quad (12)$$

The iteration (12) computes a sequence of relaxed SRSs $\{\mathcal{H}_k^\epsilon\}_{k=0}^N$. Using the formulas in Theorem 1, these sets $\{\mathcal{H}_k^\epsilon\}_{k=0}^N$ also define a sequence of "approximate" value functions and the corresponding feedback policy. Intuitively speaking, the cost associated with this relaxed policy should be very close to the optimal one if the relaxation parameter $\epsilon$ is small. In the next subsection, we shall formally analyze the performance of this policy. Before doing that, we first demonstrate the simplicity of computing such a relaxed policy.

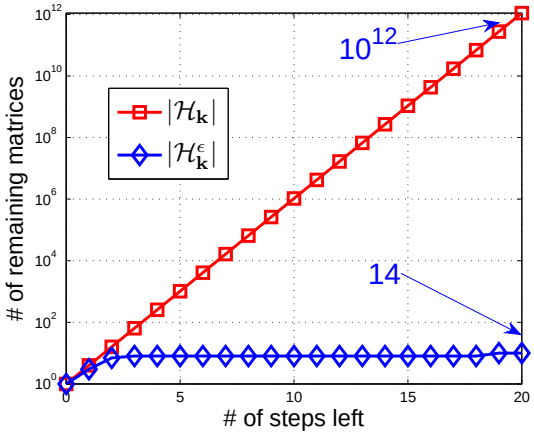


Fig. 2. Evolution of $|\mathcal{H}_k^\epsilon|$

Consider the example with the following matrices:

$$A_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 2 & 1 \\ 0 & 0.5 \end{bmatrix}, A_3 = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix},$$
$$A_4 = \begin{bmatrix} 3 & 1 \\ 0 & 0.8 \end{bmatrix}, B_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, B_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, B_3 = B_1,$$
$$B_4 = B_2, Q_i = Q_f = I_2, R_i = 1, i = 1, \ldots, 4, \text{ and } N = 20.$$

By Theorem 1, the solution of this problem is completely determined by the SRSs $\{\mathcal{H}_k\}_{k=0}^N$. As shown in Fig. 2, a direct computation of $\{\mathcal{H}_k\}_{k=0}^N$ results in a combinatorial complexity of the order $10^{12}$. However, if we use the relaxed iteration (12) with $\epsilon = 10^{-3}$, eventually $\mathcal{H}_N^\epsilon$ only contains 14 matrices. This example shows that the numerical relaxation can dramatically simplify the computation. Our next task is to prove that a small relaxation $\epsilon$ will indeed result in a suboptimal policy and will not change some important properties, such as the stability, of the close-loop system.

### B. Performance Analysis

We first introduce some notations. Let $\|\cdot\|$ be the 2-norm of a given matrix or vector. Let $\mathbb{Z}^+$ be the set of all nonnegative integers. Denote by $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ the smallest and the largest eigenvalues of a p.s.d. matrix, respectively. Define $\lambda_Q^- = \min_{i\in\mathbb{M}}\{\lambda_{min}(Q_i)\}$ and $\lambda_f^+ \triangleq \lambda_{max}(Q_f)$. For $k = 1, \ldots, N$, define $V_k^\epsilon(z) = \min_{P\in\mathcal{H}_k^\epsilon} z^T P z$ and let $\xi_k^\epsilon(\cdot)$ be the feedback law generated by $V_{k-1}^\epsilon$, namely,

$$\xi_k^\epsilon(z) = \arg\min_{(u,v)} \{L(z,u,v) + V_{k-1}^\epsilon(A_v z + B_v u)\}. \quad (13)$$

Following a similar proof of Theorem 1 ([11]), one can easily obtain:

$$\xi_k^\epsilon(z) = \Big( -K_{i_k^\epsilon(z)}(P_k^\epsilon(z)) z \,,\, i_k^\epsilon(z) \Big), \quad (14)$$
$$\text{where} \quad \Big(P_k^\epsilon(z), i_k^\epsilon(z)\Big) = \arg\min_{P\in\mathcal{H}_{k-1}^\epsilon, i\in\mathbb{M}} z^T \rho_i(P) z,$$

where $K_i(\cdot)$ denotes the Kalman gain for a given p.s.d. matrix as defined in (11). Let $\pi_N^\epsilon = \{\xi_N^\epsilon, \ldots, \xi_1^\epsilon\}$ be the $N$-horizon policy generated by $\{V_k^\epsilon\}_{k=0}^{N-1}$. Let $\pi_N^* = \{\xi_N^*, \ldots, \xi_1^*\}$ be the optimal policy generated by the *exact* value functions $\{V_k\}_{k=0}^{N-1}$. Typically, $\pi_N^\epsilon$ is much easier to

compute than $\pi_N^*$ because $\mathcal{H}_k^\epsilon$ contains much fewer matrices than $\mathcal{H}_k$. However, the relaxation $Algo_\epsilon(\cdot)$ introduces an error and this error propagates through the iteration (12). Therefore, to take advantage of the simplicity of $\pi_N^\epsilon$, it must be ensured that $J_{\pi_N^\epsilon}(z)$, namely, the actual cost associated $\pi_N^\epsilon$, does not deviate too far from the optimal cost $V_N(z)$.

The goal of this subsection is to derive conditions under which the feedback policy $\pi_N^\epsilon$ is stabilizing and suboptimal. A general $N$-horizon policy $\pi_N$ is called $\delta$-suboptimal over a set $E$ if for any initial state $x_0 \in E$, the cost under $\pi_N$ is within the $\delta$-neighborhood of the optimal cost, i.e., $|J_{\pi_N}(x_0) - V_N(x_0)| \leq \delta$. Let $x_{z,N}^*(\cdot)$ be an $N$-horizon optimal trajectory originating from $z$ at time 0. Similarly, denote by $x_{z,N}^\epsilon(\cdot)$ the $N$-horizon state trajectory driven by $\pi_N^\epsilon$ with initial condition $x_{z,k}^\epsilon(0) = z$. Define

$$\tilde{V}_{k+1}^\epsilon(z) = \min_{u,v}\{L(z,u,v) + V_k^\epsilon(A_v z + B_v u)\}. \quad (15)$$

Following immediately from (14), we have

$$\tilde{V}_{k+1}^\epsilon(z) = \min_{P\in\rho_{\mathbb{M}}(\mathcal{H}_k^\epsilon)} z^T P z.$$

The iteration (12) together with the definition of $\epsilon$-ES yields

$$\tilde{V}_{k+1}^\epsilon \leq V_{k+1}^\epsilon(z) \leq \tilde{V}_{k+1}^\epsilon(z) + \epsilon\|z\|^2. \quad (16)$$

We make the following two assumptions in the rest of this paper.

(A1) At least one subsystem is stabilizable;

(A2) $Q_i \succ 0, \forall i \in \mathbb{M}$.

*Remark 1:* The above assumptions are analogous to the stabilizability and detectability conditions commonly adopted in the study of the classical LQR problems. They are not restrictive because randomly generated subsystem matrices will satisfy them with probability 1.

*Lemma 2 ([12]):* Under assumption (A1), there exists a constant $\beta < \infty$, such that $V_N(z) \leq \beta\|z\|^2$ for all $N \in \mathbb{Z}^+$ and $z \in \mathbb{R}^n$.

Two important inequalities that are frequently used throughout the subsequent discussions are given in the following lemma.

*Lemma 3:* Under assumptions (A1) and (A2), for any integer $N \geq 0$, we have

$$V_N(z) \leq V_N^\epsilon(z) \leq V_N(z) + \epsilon\eta\|z\|^2 \quad (17)$$

$$\text{and } \tilde{V}_N^\epsilon(z) \leq V_N(z) + \epsilon(\eta-1)\|z\|^2, \quad (18)$$

where $\eta = \frac{1+(\beta/\lambda_Q^- - 1)\gamma}{1-\gamma}$.

*Proof:* See [18]. ∎

It has been proved in our previous paper [12] that under assumptions (A1) and (A2), the optimal trajectory $x_{z,N}^*(\cdot)$ is exponentially stable. Intuitively speaking, this property should also hold for $x_{z,N}^\epsilon(\cdot)$ when $\epsilon$ is sufficiently small. We now derive an upper bound of $\epsilon$ that guarantees the stability of $x_{z,N}^\epsilon(\cdot)$. The following lemma is the key in deriving this upper bound.

*Lemma 4:* Under assumptions (A1) and (A2), the trajectory $x_{z,N}^\epsilon(\cdot)$ satisfies

$$\|x_{z,N}^\epsilon(t)\|^2 \leq \left(\gamma + \frac{\epsilon\gamma\eta}{\beta}\right)^t \left(\frac{\beta+\epsilon\eta}{\lambda_Q^-}\right)\|z\|^2, \text{for } t < N,$$

$$\text{and } \|x_{z,N}^\epsilon(N)\|^2 \leq \left(\gamma + \frac{\epsilon\gamma\eta}{\beta}\right)^{N-1} \left(\frac{\zeta^2(\beta+\epsilon\eta)}{\lambda_Q^-}\right)\|z\|^2.$$

where $\beta$ is the constant defined in Lemma 2,

$$\gamma = \tfrac{1}{1+\lambda_Q^-/\beta} \text{ and } \zeta = \max_{i\in\mathbb{M}}\|A_i - B_i K_i(Q_f)\|. \quad (19)$$

*Proof:* In this proof, we denote $x_{z,N}^\epsilon(\cdot)$ by $\hat{x}(\cdot)$ and assume the corresponding hybrid control sequence is $(\hat{u}(\cdot), \hat{v}(\cdot))$. By (13), (15–18) and Lemma 2, for each $t = 1, \ldots, N$, we have

$$\begin{aligned}
&V_{N-(t-1)}^\epsilon(\hat{x}(t-1)) - V_{N-t}^\epsilon(\hat{x}(t)) \\
\geq&\tilde{V}_{N-(t-1)}^\epsilon(\hat{x}(t-1)) - V_{N-t}^\epsilon(\hat{x}(t)) \\
=&L(\hat{x}(t-1), \hat{u}(t-1), \hat{v}(t-1))) \geq \lambda_Q^-\|\hat{x}(t-1)\|^2 \\
\geq&\tfrac{\lambda_Q^-}{\beta} V_{N-(t-1)}(\hat{x}(t-1)) \\
\geq&\tfrac{\lambda_Q^-}{\beta}\left(V_{N-(t-1)}^\epsilon(\hat{x}(t-1)) - \epsilon\eta\|\hat{x}(t-1)\|^2\right) \\
\geq&\tfrac{\lambda_Q^-}{\beta} V_{N-t}^\epsilon(\hat{x}(t)) - \tfrac{\lambda_Q^-\epsilon\eta}{\beta}\|\hat{x}(t-1)\|^2.
\end{aligned}$$

Therefore, for $t = 1, \ldots, N$,

$$\begin{aligned}
&V_{N-t}^\epsilon(\hat{x}(t)) \leq \gamma\left[V_{N-(t-1)}^\epsilon(\hat{x}(t-1)) + \tfrac{\lambda_Q^-\epsilon\eta}{\beta}\|\hat{x}(t-1)\|^2\right] \\
\leq&\left[\gamma\left(1 + \tfrac{\epsilon\eta}{\beta}\right)\right] V_{N-(t-1)}^\epsilon(\hat{x}(t-1)) \\
\leq&\left(\gamma + \tfrac{\epsilon\gamma\eta}{\beta}\right)^t V_N^\epsilon(z) \leq \left(\gamma + \tfrac{\epsilon\gamma\eta}{\beta}\right)^t (\beta+\epsilon\eta)\|z\|^2.
\end{aligned}$$

Here, the second inequality follows from the fact that $V_k^\epsilon(z) \geq \lambda_Q^-\|z\|^2$ for $k \geq 0$. Using this fact again yields

$$\|\hat{x}(t)\|^2 \leq \left(\gamma + \tfrac{\epsilon\gamma\eta}{\beta}\right)^t \left(\tfrac{\beta+\epsilon\eta}{\lambda_Q^-}\right)\|z\|^2, \text{ for } t < N. \quad (20)$$

For $t = N$, according to (14), we have that $\hat{x}(N) = (A_i - B_i K_i(Q_f)) \cdot \hat{x}(N-1)$ for some $i \in \mathbb{M}$. Therefore, $\|\tilde{x}(N)\|^2 \leq \zeta^2\|\tilde{x}(N-1)\|^2$, where $\zeta$ is defined in (19). The desired result follows from (20). ∎

With Lemma 4, the following theorem follows immediately.

*Theorem 2:* If $\epsilon < \frac{(1-\gamma)\beta}{\gamma\eta}$, the policy $\pi_N^\epsilon$ is stabilizing.

We now derive an upper bound for the actual cost associated with the policy $\pi_N^\epsilon$

*Theorem 3:* Under assumptions (A1) and (A2), $J_{\pi_N^\epsilon}(z) \leq V_N(z) + \epsilon(\eta-1)\|z\|^2$. for any $z \in \mathbb{R}^n$ and $N \geq 0$.

*Proof:* Let $\hat{x}(\cdot)$ and $(\hat{u}(\cdot), \hat{v}(\cdot))$ be the same as in the proof of Lemma 4. By (13) and (15), we have $L(\hat{x}(t), \hat{u}(t), \hat{v}(t)) = \tilde{V}_{N-t}^\epsilon(\hat{x}(t)) - V_{N-(t+1)}^\epsilon(\hat{x}(t+1))$ for

each $t = 0, \ldots, N-1$. Therefore,

$$\begin{aligned} J_{\pi_N^\epsilon}(z) &= \sum_{t=0}^{N-1} L(\hat{x}(t), \hat{u}(t), \hat{v}(t)) + \psi(\hat{x}(N)) \\ &= \sum_{t=0}^{N-1} [\tilde{V}_{N-t}^\epsilon(\hat{x}(t)) - V_{N-(t+1)}^\epsilon(\hat{x}(t+1))] + \psi(\hat{x}(N)) \\ &= \tilde{V}_N^\epsilon(z) + \sum_{t=1}^{N-1} \Big\{ [\tilde{V}_{N-t}^\epsilon(\hat{x}(t)) - V_{N-t}^\epsilon(\hat{x}(t))] \\ &\qquad + [\psi(\hat{x}(N)) - V_0^\epsilon(\hat{x}(N))] \Big\}. \end{aligned}$$

Since by definition $\psi(z) = V_0^\epsilon(z)$ and $\tilde{V}_{N-t}^\epsilon(z) \leq V_{N-t}^\epsilon(z)$ for any $z \in \mathbb{R}^n$ and $t = 1, \ldots N-1$, we have

$$J_{\pi_N^\epsilon}(z) \leq \tilde{V}_N^\epsilon(z) \leq V_N(z) + \epsilon(\eta - 1)\|z\|^2.$$

■

*Remark 2:* Notice that the error function $\epsilon(\eta - 1)\|z\|^2$ does not grow with respect to the horizon $N$. This property plays a crucial role in deriving the suboptimal policies for the infinite-horizon DSLQR problems.

*Corollary 1:* Under the same conditions as in Theorem 3, $\pi_N^\epsilon$ is $\delta$-suboptimal over the unit ball if $\epsilon \leq \frac{\delta}{\eta - 1}$.

Based on our analysis, a $\delta$-suboptimal policy can be obtained using the following Algorithm.

**Algorithm 2** (Suboptimal Control in Finite Horizon)

1) **Initialization:** Specify an error tolerance $\delta$. Let $\epsilon = \frac{\delta}{\eta - 1}$ and set $\mathcal{H}_0^\epsilon = Q_f$
2) **Relaxed ES Iteration:** Perform iteration (12) over the whole horizon $N$.
3) **Suboptimal Strategy:** The suboptimal $N$-horizon policy $\pi_N^\epsilon = \{\xi_N^\epsilon(x), \ldots, \xi_1^\epsilon(x)\}$ is given by:

$$\xi_k^\epsilon(x) = \Big( -K_{i_k^\epsilon(x)}(P_k^\epsilon(x))x \,,\, i_k^\epsilon(x) \Big),$$

$$\text{where} \quad \Big( P_k^\epsilon(x), i_k^\epsilon(x) \Big) = \underset{P \in \mathcal{H}_{k-1}^\epsilon, i \in \mathbb{M}}{\arg\min} \; x^T \rho_i(P) x.$$

## V. Extension to Large or Infinite Horizon

It is natural to solve the infinite-horizon case in a divide-and-conquer manner, namely, by applying Algorithm 2 to a reasonably large size of subhorizon, $m$, and then extending the obtained optimal strategy periodically. We now show that, by choosing proper $m$ and $\epsilon$, such a periodic policy can achieve an arbitrary suboptimal performance. Let $\hat{\pi}_m^\epsilon = (\hat{\xi}_m^\epsilon, \ldots, \hat{\xi}_1^\epsilon)$ be the $m$-horizon policy returned by Algorithm 2 with $Q_f = 0$. It follows from Theorem 3 that

$$J_{\hat{\pi}_m^\epsilon}(z) \leq V_m^0(z) + \epsilon(\eta - 1)\|z\|^2 \leq V^*(z) + \epsilon(\eta - 1)\|z\|^2, \quad (21)$$

where $V_m^0(z)$ denotes the $m$-horizon value function with $Q_f = 0$. For $m \geq 2$, let $\pi_\infty^{\epsilon,m}$ be the periodic extension of the first $m-1$ terms of $\hat{\pi}_m^\epsilon$, i.e.[1],

$$\pi_\infty^{\epsilon,m} = \{\hat{\xi}_m^\epsilon, \ldots, \hat{\xi}_2^\epsilon, \hat{\xi}_m^\epsilon, \ldots, \hat{\xi}_2^\epsilon, \ldots\}. \quad (22)$$

[1]As can be seen from Lemma 4, by using only the first $m-1$ terms of $\hat{\pi}_m^\epsilon$ in constructing $\pi_\infty^{\epsilon,m}$, we can obtain a better bound for the convergence of the close-loop trajectory.

We first establish conditions under which the specially constructed policy $\pi_\infty^{\epsilon,m}$ is stabilizing.

*Theorem 4:* Under assumptions (A1) and (A2), if $\epsilon < \frac{(1-\gamma)\beta}{\gamma\eta}$ and $m > \frac{\ln \lambda_Q^- - \ln(\beta + \epsilon\eta)}{\ln(\beta\gamma + \epsilon\gamma\eta) - \ln \beta} + 1$, then $\pi_\infty^{\epsilon,m}$ is exponentially stabilizing.

*Proof:* Denote by $\hat{x}(\cdot)$ the trajectory associated with the policy $\pi_\infty^{\epsilon,m}$ with initial condition $\hat{x}(0) = z$. Let

$$c_m = \left(\gamma + \frac{\epsilon\gamma\eta}{\beta}\right)^{m-1} \left(\frac{\beta + \epsilon\eta}{\lambda_Q^-}\right). \quad (23)$$

It can be easily verified that under our assumptions, $c_m$ is strictly smaller than 1. By Lemma (4), we have $\|\hat{x}(k(m-1))\|^2 \leq c_m \|\hat{x}((k-1)(m-1))\|^2$ for all $k \geq 1$. Thus, $\|\hat{x}(\cdot)\|^2$ must decrease by a factor of $c_m < 1$ in every $m-1$ steps. It follows that the policy $\pi_\infty^{\epsilon,m}$ is exponentially stabilizing. ■

We now derive a bound for the error between the actual cost $J_{\pi_\infty^{\epsilon,m}}(z)$ and the optimal cost $V^*(z)$.

*Theorem 5:* Under the same conditions as in Theorem 4, we have

$$V^*(z) \leq J_{\pi_\infty^{\epsilon,m}}(z) \leq V^*(z) + \frac{c_m\beta + \epsilon(\eta - 1)}{1 - c_m}\|z\|^2, \quad (24)$$

where $c_m$ is defined in (23).

*Proof:* Obviously, $V^*(z) \leq J_{\pi_\infty^{\epsilon,m}}(z)$ as $\pi_\infty^{\epsilon,m}$ is an infinite-horizon policy. Let $\hat{x}(\cdot)$ be the system trajectory generated by the policy $\pi_\infty^{\epsilon,m}$ starting from $\hat{x}(0) = z$. Define $z_i = \hat{x}(i \cdot (m-1))$ for $i = 0, 1, \ldots$. Let $\tilde{\pi} \triangleq \{\hat{\xi}_m^\epsilon, \ldots, \hat{\xi}_2^\epsilon\}$ be the first $m-1$ terms of $\hat{\pi}_m^\epsilon$. Then by (21),

$$\begin{aligned} J_{\pi_\infty^{\epsilon,m}}(z) &= \sum_{i=0}^{\infty} J_{\tilde{\pi}}(z_i) \leq \sum_{i=0}^{\infty} J_{\hat{\pi}_m^\epsilon}(z_i) \\ &\leq \sum_{i=0}^{\infty} [V^*(z_i) + \epsilon(\eta - 1)\|z_i\|^2]. \end{aligned}$$

By Lemma (4), $\|z_i\|^2 \leq c_m^i \|z\|^2$, where $c_m < 1$ is defined in (23). Therefore, $J_{\pi_\infty^{\epsilon,m}}(z) \leq V^*(z) + \frac{c_m\beta + \epsilon(\eta-1)}{1 - c_m}\|z\|^2$ for any initial state $z$. ■

With the above result, we can easily derive a lower bound for $m$ that guarantees the $\delta$-suboptimality of $\pi_\infty^{\epsilon,m}$ for an arbitrary $\delta > 0$.

*Corollary 2:* Suppose that the conditions in Theorem 4 hold. For any $\delta > 0$, if we further have $\epsilon < \frac{\delta}{\eta - 1}$ and

$$m > m_\infty^{\delta,\epsilon} \triangleq \frac{\ln[\delta - \epsilon(\eta-1)]\lambda_Q^- - \ln(\beta + \delta)(\beta + \epsilon\eta)}{\ln(\beta\gamma + \epsilon\gamma\eta) - \ln \beta} + 1, \quad (25)$$

then the policy $\pi_\infty^{\epsilon,m}$ is $\delta$-suboptimal over the unit ball.

For a given tolerance $\delta$ on the optimal cost, we only need to perform $m_\infty^{\delta,\epsilon}$ steps of the approximate value iterations (12). The obtained value functions $\{V_k^\epsilon(z)\}_{k=0}^{m_\infty^{\delta,\epsilon}}$ characterize the $m_\infty^{\delta,\epsilon}$-horizon feedback policy $\hat{\pi}_m^\epsilon$ which can be used periodically to construct an infinite-horizon policy $\pi_\infty^{\epsilon,m}$. By Corollary 2, such a periodic policy is guaranteed to be $\delta$-suboptimal within the unit ball. This idea also applies when the horizon is large but finite. Denote by $[\pi_\infty^{\epsilon,m}]_N$ the $N$-horizon truncation of the policy $\pi_\infty^{\epsilon,m}$, i.e., $[\pi_\infty^{\epsilon,m}]_N(t) = \pi_\infty^{\epsilon,m}(t)$ for $t = 0, \ldots, N-1$. Similar performance bound as in Theorem 5 can be derived for $[\pi_\infty^{\epsilon,m}]_N(t)$ .

*Theorem 6:* Under the same conditions as in Theorem 4, for any $N \geq m$, we have

$$J_{[\pi_\infty^{\epsilon,m}]_N}(z) \leq V_N(z) + \left[\frac{c_m\beta + \epsilon(\eta-1)}{1-c_m} + \lambda_f^+ c_m^{N_m}\right]\|z\|^2,$$

where $c_m$ is defined in (23) and $N_m = \lfloor N/(m-1) \rfloor$.

*Proof:* Denote by $\hat{x}(\cdot)$ the close-loop trajectory generated by the policy $[\pi_\infty^{\epsilon,m}]_N$. Let $\tilde{\pi}$ and $z_i$ be the same as in the proof of Theorem 5. Then by (21),

$$J_{[\pi_\infty^{\epsilon,m}]_N}(z) - \psi(\hat{x}(N)) \leq \sum_{i=0}^{N_m+1} J_{\tilde{\pi}}(z_i)$$
$$\leq \sum_{i=0}^{N_m+1} J_{\hat{\pi}_m^\epsilon}(z_i) \leq \sum_{i=0}^{N_m+1} \left[V_m^0(z_i) + \epsilon(\eta-1)\|z_i\|^2\right]$$

Notice that $V_m^0(z) \leq V_N(z)$, $V_m^0(z_i) \leq V^*(z_i)$ and $V^*(z_i) \leq \beta\|z_i\|^2 \leq \beta c_m^i \|z_i\|^2$, by adding some small positive terms, we have

$$J_{[\pi_\infty^{\epsilon,m}]_N}(z) - \psi(\hat{x}(N))$$
$$\leq V_N(z) + \sum_{i=1}^{\infty} \beta c_m^i \|z\|^2 + \sum_{i=0}^{\infty} \epsilon(\eta-1) c_m \|z\|^2. \quad (26)$$

By our hypotheses, we have $c_m < 1$. Thus, $J_{[\pi_\infty^{\epsilon,m}]_N}(z) - \psi(\hat{x}(N)) \leq V_N(z) + \frac{c_m\beta+\epsilon(\eta-1)}{1-c_m}\|z\|^2$. Considering $\psi(\hat{x}(N)) \leq \lambda_f^+\|\hat{x}(N)\|^2 \leq \lambda_f^+ c_m^{N_m}\|z\|^2$, the desired result is proved. ■

*Corollary 3:* Suppose the conditions in Theorem 4 hold. For any $\delta > 0$, if we further have $\epsilon < \frac{\delta}{\eta-1}$ and

$$N \geq m > m_N^{\delta,\epsilon} \triangleq \frac{\ln[\delta - \epsilon(\eta-1)]\lambda_Q^- - \ln(\beta+\delta+\sigma_f^+)(\beta+\epsilon\eta)}{\ln(\beta\gamma+\epsilon\gamma\eta) - \ln\beta} + 1, \quad (27)$$

then the $N$-horizon policy $[\pi_\infty^{\epsilon,m}]_N$ is $\delta$-suboptimal over the unit all.

Motivated by the above analysis, for a large or infinite $N$, a $\delta$-suboptimal $N$-horizon policy can be obtained as follows. First, find the largest $\epsilon$ that satisfies all the conditions in Corollary 2. Second, let $m = m_\infty^{\delta,\epsilon}$ or $m = m_N^{\delta,\epsilon}$ depending on whether $N$ is infinite or not. Third, compute the $m$-horizon suboptimal policy $\hat{\pi}_m^\epsilon$ using Algorithm 2 with $Q_f = 0$. Finally, use $\hat{\pi}_m^\epsilon$ to construct $\pi_\infty^{\epsilon,m}$ based on (22) and keep the first $N$ steps of $\pi_\infty^{\epsilon,m}$ to obtain an $N$-horizon policy $[\pi_\infty^{\epsilon,m}]_N$[2]. By Corollary 2 or 3, $[\pi_\infty^{\epsilon,m}]_N$ is guaranteed to be $\delta$-suboptimal over the unit ball. The above procedure of constructing the suboptimal control policy is summarized in Algorithm 3. Note that in this procedure, we have assumed that $N > m$. If this is not the case, we should still use Algorithm 2 to carry out the approximate iterations (12) for the whole horizon $N$.

*Remark 3:* The analytical bounds $m_\infty^{\delta,\epsilon}$ and $m_N^{\delta,\epsilon}$ derived in (25) and (27) may be conservative for some applications. An alternative approach is to start from a smaller value for $m$ in Step 2) of Algorithm 3 and gradually increase its value until the performance saturates. Our analysis guarantees that this tentative procedure can eventually reach any prespecified suboptimal performance by gradually increasing $m$.

[2]If $N$ is infinite, the policy $[\pi_\infty^{\epsilon,m}]_N$ would be the same as $\pi_\infty^{\epsilon,m}$

**Algorithm 3** (Large or infinite Horizon Suboptimal Control)

1) **Initialization:** Specify an error tolerance $\delta$. Let $\epsilon = \max\{\frac{\delta}{\eta-1}, \frac{\beta(1-\gamma)}{\gamma\eta}\}$.
2) **# of iterations steps:** If $N = \infty$, let $m = m_\infty^{\delta,\epsilon}$; otherwise, let $m = m_N^{\delta,\epsilon}$. If $N \leq m$, stop and turn to Algorithm 2.
3) **m-horizon Policy:** Calculate the $m$-horizon suboptimal policy $\hat{\pi}_m^\epsilon$ using Algorithm 2 with $Q_f = 0$.
4) **Horizon Extension:** Construct $\pi_\infty^{\epsilon,m}$ from $\hat{\pi}_m^\epsilon$ using (22) and keep its first $N$ terms to obtain $[\pi_m^\infty]_N$.

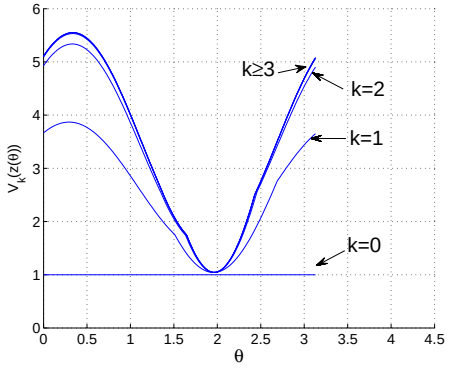


Fig. 3. Convergence of the Value function of Example VI-A.

## VI. Numerical Examples

### A. Example 1

First consider a simple DSLQR problem with control horizon $N = 1000$ and two second-order subsystems:

$$A_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, A_2 = \begin{bmatrix} 2 & 1 \\ 0 & 0.5 \end{bmatrix}, B_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Suppose that the state and control weights are $Q_1 = Q_2 = I_2$ and $R_1 = R_2 = 1$, respectively. Both subsystems are unstable but controllable. Algorithm 3 is applied to solve this DSLQR problem. With $\delta = 10^{-3}$, the upper bound of the required number of iterations is $m_N^{\delta,\epsilon} = 56$, while as observed in the simulation, the value function already converges in 6 steps. Since $V_k(z)$ is homogeneous and symmetric, in Fig. 3, we plot the evolution of the value functions on the upper half of the unit circle, i.e. the points of the form $z(\theta) = [\cos(\theta), \sin(\theta)]^T$ with $\theta \in [0, \pi]$. The number of matrices in $\mathcal{H}_k^\epsilon$ at each step $k$ is listed in Table I. It can be seen that $|\mathcal{H}_k^\epsilon|$ is indeed very small and stays at the maximum value 5 as opposed to growing exponentially as k increases.

TABLE I
$|\mathcal{H}_k^\epsilon|$ for Example VI-A

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\lvert\mathcal{H}_k^\epsilon\rvert$ | 2 | 4 | 5 | 5 | 5 | 5 |

### B. Random Examples

To further demonstrate its effectiveness, Algorithm 3 is tested by two sets of randomly generated DSLQR problems.

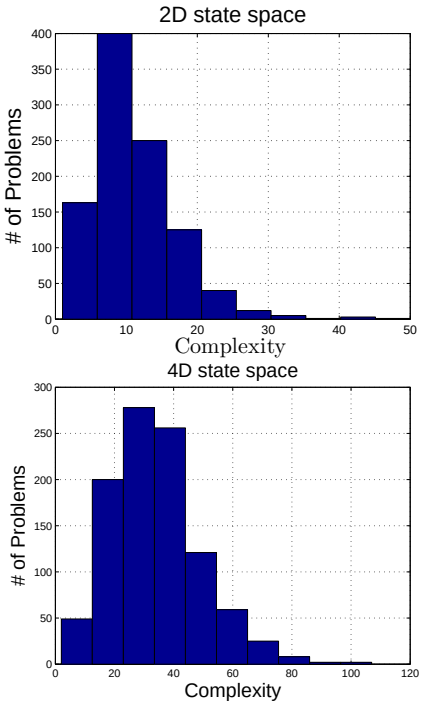


Fig. 4. Complexity distributions of the random examples.

The first set consists of 1000 two-dimensional DSLQR problems with 10 subsystems. The second set consists of 1000 four-dimensional DSLQR problems with 4 subsystems. For both sets, the control horizon $N$ is infinite and $\delta = 10^{-3}$. All of these problems are successfully solved by Algorithm 3 and the distributions of the complexity, namely, the maximum numbers of matrices required for characterizing the suboptimal policy, are plotted in Fig. 4. It can be seen from the figure that all of the two-dimensional problems require less than 50 matrices and a majority of them only need less than 15 matrices. However, a majority of the four-dimensional problems need about 40 matrices and some of them may need more than 100 matrices. The complexity of Algorithm 3 depends heavily on the state dimension. In a higher dimensional state space, a larger $\delta$ is usually needed in order to retain a high computational speed.

## VII. Conclusion

The value functions and the optimal strategies of the DSLQR problem can be exactly characterized by the SRSs $\mathcal{H}_k$, whose size grows exponentially fast. However, with some small relaxation, many matrices in $\mathcal{H}_k$ becomes redundant in terms of characterizing the suboptimal strategies. Efficient algorithms are developed to prune out these redundant matrices and to compute the suboptimal strategies. Analytical conditions under which the strategies generated by the algorithms are stabilizing and suboptimal are derived explicitly in terms of subsystem matrices. The proposed algorithm and the analysis provide a systematical way of solving the finite-horizon and infinite-horizon DSLQR problem with guaranteed close-loop stability and suboptimality. The results of this paper can be used to study many other problems of the switched linear systems, such as the switched Kalman filtering problem, the switched LQG problem, and the switched receding horizon control problem, etc. All of these will be our future research directions.

## References

[1] W. Zhang and J. Hu. Dynamic buffer management using optimal control of hybrid systems. *Automatica*, (7):1831–1840, Jul 2008.

[2] C. G. Cassandras, D. L. Pepyne, and Y. Wardi. Optimal control of a class of hybrid systems. *IEEE Transactions on Automatic Control*, 46(3):398–415, 2001.

[3] X. Xu and P.J. Antsaklis. Optimal control of switched systems based on parameterization of the switching instants. *IEEE Transactions on Automatic Control*, 49(1):2–16, 2004.

[4] M. Egerstedt, Y. Wardi, and F. Delmotte. Optimal control of switching times in switched dynamical systems. *IEEE Transactions on Automatic Control*, 51(1):110–115, 2006.

[5] C. Seatzu, D. Corona, A. Giua, and A. Bemporad. Optimal control of continuous-time switched affine systems. *IEEE Transactions on Automatic Control*, 51:726– 741, May 2006.

[6] Sven Hedlund and Anders Rantzer. Convex dynamic programming for hybrid systems. *IEEE Transactions on Automatic Control*, 47(9):1536–1540, sep 2002.

[7] S. Hedlund and A. Rantzer. Optimal control of hybrid system. In *Proceedings of the IEEE Conference on Decision and Control*, volume 4, pages 3972–3977, Phoenix, AZ, December 1999.

[8] A. Rantzer and M. Johansson. Piecewise linear quadratic optimal control. *IEEE Transactions on Automatic Control*, Apr. 2000.

[9] A. Bemporad, A. Giua, and C. Seatzu. Synthesis of state-feedback optimal controllers for continuous-time switched linear systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 3182– 3187, 2002.

[10] F. Borrelli, M. Baotic, A. Bemporad, and M. Morari. Dynamic programming for constrained optimal control of discrete-time linear hybrid systems. *Automatica*, 41:1709–1721, October 2005.

[11] Wei Zhang and Jianghai Hu. On Optimal Quadratic Regulation for Discrete-Time Switched Linear Systems. In *International Workshop on Hybrid Systems: Computation and Control*, pages 584–597, St Louis,MO,USA, 2008.

[12] Wei Zhang and Jianghai Hu. On the value functions of the optimal quadratic regulation problem for discrete-time switched linear systems. In *IEEE Conference on Decision and Control*, Cancun, Mexico, December 2008.

[13] D. P. Bertsekas and J. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, Sep. 1996.

[14] B. Lincoln and A. Rantzer. Relaxing dynamic programming. *IEEE Transactions on Automatic Control*, 51(8):1249–1260, August 2006.

[15] D.P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 2. Athena Scientific, 2 edition, 2001.

[16] Wei Zhang and Jianghai Hu. Optimal quadratic regulation for discrete-time switched linear systems: a numerical approach. In *IEEE Proceedings of the American Control Conference*, Seattle, WA, Jun 2008.

[17] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[18] W. Zhang and J. Hu. Switched lqr problem in discrete time. submitted for journal publication (available upon request).