Escaping Locally Optimal Decentralized Control Polices via Damping

Han Feng and Javad Lavaei

Industrial Engineering and Operations Research, University of California, Berkeley

***Abstract*— We study the design of an optimal static decentralized controller with a quadratic cost and propose a variant of the homotopy continuation method using a damping technique. This method generates a series of optimal distributed control (ODC) problems via a continuous variation of the system parameters. Diverging from the classical theme of a tracking a specific trajectory of locally optimal controllers for these ODC problems, we focus on the possibility of leveraging local-search algorithms to locate among several locally optimal controller trajectories the globally optimal trajectory. We analyze the continuity and asymptotic properties of the locally and globally optimal controller trajectories as the damping parameter varies. In particular, we prove that under certain conditions, there is no spurious locally optimal controller for an ODC problem with favorable control structure and a large damping parameter. As a result, the proposed method is able to locate the globally optimal trajectory with a suitable discretization in the space of the damping parameter. To demonstrate the effectiveness of this technique, it is shown that even for instances with an exponential number of connected components, damping could merge the trajectories of all local solutions to the trajectory of the global solutions. We further illustrate the convoluted behavior of the locally optimal trajectories with numerical examples on random systems.**

I. Introduction

The optimal decentralized control (ODC) problem adds controller constraints to the classical centralized optimal control problem. This addition breaks down the separation principle and the classical solution formulas culminated in [3]. Although ODC has been proved intractable in general [1], [19], the problem has convex formulations under various assumptions such as partially nestedness [16], positiveness [14], and quadratic invariance [10]. More recently, the System Level Framework [18] identifies a large set of problems that have a convex formulation. However, it is challenging to solve large-scale optimization problem arising from the convex relaxations and reformulations.

As an alternative to convexification techniques with a high computational complexity, local search methods are extensively used in the practice of optimization. This approach stands out for many problems in machine learning, where it is empirically and theoretically shown that simple policy search methods with stochastic gradient descent are able to effectively solve non-convex optimization or learning problems in practical scenarios [5], [8], [9]. Many efficiency statements of local search from the machine learning literature, however, are unlikely to directly carry over to ODC, due to the recent investigation of the topological properties of ODC in [6] showing that — unlike many problems in machine learning — ODC can have an exponential number of locally optimal solutions, and therefore, the landscape of optimization is highly complex.

This paper attempts to delineate the boundary of tractable ODC instances that are solvable by local-search methods, by studying the evolution of locally optimal decentralized controllers as the system dynamics vary. We have recently proved that one variation of the system dynamics called "damping" effectively reduces the topological complexity of the set of stabilizing decentralized controllers [6]. The main objective of the present paper is to show how damping reduces the number of locally optimal decentralized controllers. We prove continuity and asymptotic properties of the trajectories of the locally optimal solutions. Notably, the analysis leads to the result that if the system dynamics is dampened enough, there is no spurious locally optimal controller, by which we mean all locally optimal controllers are globally optimal for the damped system. The damped system, therefore, is a tractable approximate ODC problem. Furthermore, we show that this globally optimal controller in the damped system can be continuously connected to the globally optimal controller in the original system, if the globally optimal decentralized controllers are unique in the damping process.

This work is closely related to continuation methods such as homotopy. They are known to be appealing yet theoretically poorly understood [12]. Homotopy has been used as an initialization strategy in optimal control: in [2], the author mentioned the idea of gradually moving from a stable system to the original system to obtain a stabilizing controller. The paper [20] considered the H_2-reduced order problem and proposed several homotopy maps and initialization strategies; in its numerical experiments, initialization with a large multiple of $-I$ was found appealing. The paper [4] compared descent and continuation algorithms for the H_2 optimal reduced-order control problem and concluded that homotopy methods are empirically superior to descent methods. The difficulty of obtaining a convergence theory for a general constrained optimal control problem can be appreciated from the examples in [11]. Compared with those earlier works, we consider a special type of continuation named damping, to improve the locally optimal solutions in optimal decentralized control. We de-emphasize the problem

Email: han_feng@berkeley.edu, lavaei@berkeley.edu
This work was supported by grants from ARO, ONR, AFOSR, and NSF.

of accurately following a given path and instead focus on the evolution of several paths as well as the movement of locally optimal solutions from one path to another.

The remainder of this paper is organized as follows. Notations and problem formulations are given in Section II. Continuity and asymptotic properties of the damping strategy are outlined in Section III and Section IV, respectively. Numerical experiments are detailed in Section V. Concluding remarks are drawn in Section VI.

II. Problem Formulation

We study the optimal decentralized control (ODC) problem with a static controller and a quadratic cost. Consider the linear time-invariant system

$$\dot{x}(t) = Ax(t) + Bu(t),$$

where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ are real matrices of compatible sizes. The vector $x(t)$ is the state of the system with an unknown initialization $x(0) = x_0$, where x_0 is modeled as a random variable with zero mean and a positive-definite covariance $\mathbb{E}[x(0)x(0)^\top] = D_0$ (note that $\mathbb{E}[\cdot]$ denotes the expectation operator). The control input $u(t)$ is to be determined via a static state-feedback law $u(t) = Kx(t)$ with the gain $K \in \mathbb{R}^{m\times n}$ such that some quadratic performance measure is maximized. Given a controller K, the closed-loop system is

$$\dot{x}(t) = (A + BK)x(t).$$

A matrix is said to be stable if all its eigenvalues lie in the open left-half of the complex plane. The controller K is said to stabilize the system (A, B) if $A+BK$ is stable. ODC optimizes over the set of structured stabilizing controllers

$$\mathcal{K}_S = \{K : A + BK \text{ is stable}, K \in \mathcal{S}\}, \tag{1}$$

where $\mathcal{S} \subseteq \mathbb{R}^{m\times n}$ is a linear subspace of matrices, often specified by fixing certain entries of the matrix to zero. In that case, the sparsity pattern can be equivalently described with the indicator matrix $I_\mathcal{S}$, whose (i, j)-entry is defined to be

$$[I_\mathcal{S}]_{ij} = \begin{cases} 1, & \text{if } K_{ij} \text{ is free} \\ 0, & \text{if } K_{ij} = 0. \end{cases}$$

The structural constraint $K \in \mathcal{S}$ is then equivalent to $K \circ I_S = K$, where $\circ$ denotes entry-wise multiplication. In the following, we will consider a sequence of damped cost functions, which are defined as

$$\begin{aligned} J(K,\alpha) =& \mathbb{E}\int_0^\infty \left[e^{-2\alpha t}\left(\hat{x}^\top(t)Q\hat{x}(t) + \hat{u}^\top(t)R\hat{u}(t)\right)\right]dt \\ s.t. \quad & \dot{\hat{x}}(t) = A\hat{x}(t) + B\hat{u}(t) \\ & \hat{u}(t) = K\hat{x}(t). \end{aligned} \tag{2}$$

where $Q \succeq 0$ is positive semi-definite and $R \succ 0$ is positive-definite. The expectation is taken over x_0. Setting $x(t) = e^{-\alpha t}\hat{x}(t)$ and $u(t) = e^{-\alpha t}\hat{u}(t)$, the cost $J(K, \alpha)$ can be equivalently written as

$$\begin{aligned} J(K,\alpha) =& \mathbb{E}\int_0^\infty \left[x^\top(t)Qx(t) + u^\top(t)Ru(t)\right]dt \\ s.t. \quad & \dot{x}(t) = (A - \alpha I)x(t) + Bu(t) \\ & u(t) = Kx(t), \end{aligned} \tag{3}$$

ODC is commonly defined for $\alpha = 0$ as optimizing (3) over the set of stabilizing structured controllers (1). Formally

$$\begin{aligned} \min_K \quad & J(K, 0) \\ s.t. \quad & K \text{ stabilizes } (A, B) \\ & K \in \mathcal{S}. \end{aligned}$$

In our setting, the notion of stability is relaxed for a positive α. We define K as a stabilizing solution to (3) if K stabilizes the system $(A-\alpha I, B)$, in which case formulation (2) is also meaningful. Formally, we define ODC with damping as

$$\begin{aligned} \min_K \quad & J(K, \alpha) \\ s.t. \quad & K \text{ stabilizes } (A - \alpha I, B) \\ & K \in \mathcal{S}. \end{aligned} \tag{4}$$

Our relaxed notion of stability coincides with ODC when $\alpha = 0$. We emphasize that the relaxation of stability in the damped regime is a solution method, while the ultimate goal is to obtain an optimal stabilizing controller for the undamped system with $\alpha = 0$. We shall denote the problem (4) by ODC(α). We write ODC(α, K_0) if a stabilizing controller K_0 is given (to be used for the initialization of local search methods).

The two equivalent formulations above motivate the notion of "damping property". We make a formal statement below.

Lemma 1: The function $J(K, \alpha)$ defined in (2) and (3) satisfies the following "damping property": suppose that K stabilizes the system $(A - \alpha I, B)$, then for all $\beta > \alpha$, K stabilizes the system $(A - \beta I, B)$ and satisfies the relation $J(K, \beta) < J(K, \alpha)$.

Proof: By formulation (4), when $A-\alpha I+BK$ is stable and $\beta > \alpha$, it holds that $A - \beta I + BK = (A - \alpha I + BK) - (\beta - \alpha)I$ is stable. Therefore, $J(K, \beta)$ is well-defined. By formulation (2), we have $J(K, \beta) < J(K, \alpha)$. ■

We denote the set of globally optimal controllers of the damped ODC problem (4) by $K^*(\alpha)$, and the set of locally optimal controllers by $K^\dagger(\alpha)$. The set $K^\dagger(\alpha)$ contains those controllers K that satisfy the following first-order optimality conditions (see [15] for their derivation):

$$\begin{aligned} &(A-\alpha I + BK)^\top P_\alpha(K) + \\ &\quad P_\alpha(K)(A - \alpha I + BK) + K^\top RK + Q = 0 \end{aligned} \tag{5a}$$

$$\begin{aligned} &L_\alpha(K)(A-\alpha I + BK)^\top + \\ &\quad (A - \alpha I + BK)L_\alpha(K) + D_0 = 0 \end{aligned} \tag{5b}$$

$$\left[(B^\top P_\alpha(K) + RK)L_\alpha(K)\right] \circ I_S = 0 \tag{5c}$$

$$K \circ I_S = K. \tag{5d}$$

The matrices $P_\alpha(K)$ and $L_\alpha(K)$ are the closed-loop Gramians. The above conditions provide a closed-form expression for the cost

$$J(K,\alpha) = \mathrm{tr}(D_0 P_\alpha(K)), \tag{6}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Given α, the equations (5a)-(5d) and (6) are algebraic, involving only polynomial functions of the unknown matrices K, P_α and L_α. The matrices P_α and L_α are written as a function of K because they are uniquely determined from (5a) and (5b) given a stabilizing controller K. When the context is clear, we drop the implicit dependence on K in the notations P_α and L_α.

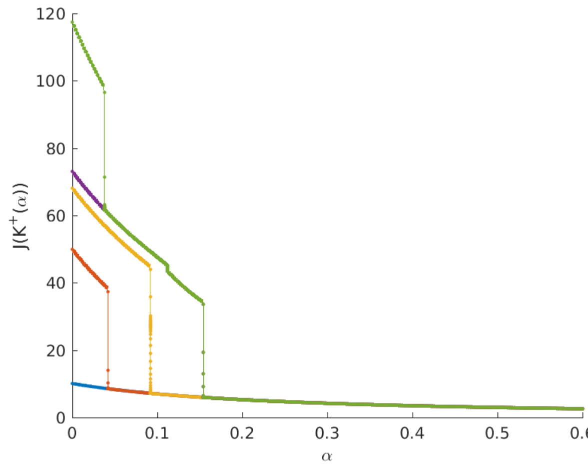
The paper studies the properties of $K^*(\alpha)$, $K^\dagger(\alpha)$, and $J(K,\alpha)$ for any control K belonging to $K^*(\alpha)$ or $K^\dagger(\alpha)$. To motivate the study of $K^\dagger(\alpha)$, Figure 1 illustrates the evolution of five locally optimal distributed controllers for a particular system as α varies (see Section V for details on the experiment). It is known that systems of this type have a large number of locally optimal controllers [6]. Figure 1a plots selected trajectories of $J(K,\alpha)$ against α, where $K \in K^\dagger(\alpha)$. The selected trajectories are connected to a stabilizing controller in $K^\dagger(0)$. The lowest curve corresponds to $J(K^*(\alpha),\alpha)$. Figure 1b plots the distance of the selected $K \in K^\dagger(\alpha)$ from the controller $K \in K^*(\alpha)$.

The fact that even modest damping causes the locally optimal trajectories to "collapse" to each other is an attractive phenomenon. Especially, this leads to the following two strategies for solving the ODC problem, which are detailed in Algorithm 1 and Algorithm 2 below.

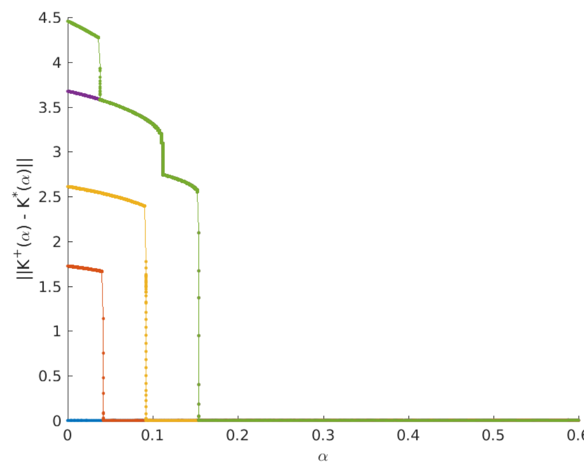
Algorithm 1 Improving an Existing Stabilizing Controller: The Forward-Backward Method

Input: $J(K,\alpha)$ and $K_0 \in S$ that stabilizes the system (A,B).
Output: A potentially improved $K_0 \in K^\dagger(0)$.
Select a list of parameters $0 = \alpha_0 < \alpha_1, \ldots, < \alpha_T$.
for $t \leftarrow 1, \ldots, T$ **do**
  Obtain a $K_t \in K^\dagger(\alpha_t)$ by solving ODC(α_t, K_{t-1}) using local search.
end for
for $t \leftarrow T-1, T-2, \ldots, 0$ **do**
  Obtain a $K_t \in K^\dagger(\alpha_t)$ by solving ODC(α_t, K_{t+1}) using local search.
end for

Algorithm 2 shall avoid many unnecessary local optimum. It starts with a large enough α for which $K = 0$ is an initial stabilizing controller in the set $\mathcal{S}$ and iteratively solves for a better controller while reducing the damping parameter α. The improvement at $\alpha = \alpha_t$ is achieved using local-search and the initialization K_{t+1} from the previous step. Algorithm 1 has the potential to improve the locally optimal controllers obtained from any method. It is different from Algorithm 2 in that it starts with a potentially undesirable controller for $\alpha = 0$ and gradually increases α to obtain an improved optimal controller for a highly-damped system



(a) Locally optimal cost trajectories against the damping parameter



(b) Distance between $K^\dagger(\alpha)$ and $K^*(\alpha)$

Fig. 1. Samples of locally optimal cost and locally optimal controller trajectories of system given in equation (15) as the damping parameter α varies.

Algorithm 2 Obtain a Stabilizing Controller: The Backward Method

Input: $J(K,\alpha)$
Output: A potentially stabilizing $K_0 \in K^\dagger(0)$.
Select a list of parameters $0 = \alpha_0 < \alpha_1, \ldots, < \alpha_T$, where α_T is large enough such that $K_T = 0$ stabilizes the system $(A - \alpha_T I, B)$.
for $t \leftarrow T-1, T-2, \ldots, 0$ **do**
  Obtain a $K_t \in K^\dagger(\alpha_t)$ by solving ODC(α_t, K_{t+1}) using local search.
end for

and then applies a variant of Algorithm 2 to backtrack that controller to a globally optimal controller for $\alpha = 0$.

Due to the NP-hardness of ODC, one cannot expect any guarantee for producing a globally optimal, or even a stabilizing, decentralized controller, unless certain conditions are met, which will be discussed later. The breakdown of these strategies will be discussed in Section V.

III. Continuity

This section studies the continuity properties of $K^*(\alpha)$ and $K^\dagger(\alpha)$. The key notion of hemi-continuity captures the evolution of parametrized optimization problems. The reader is referred to [13] for an accessible treatment.

Definition 1: The set valued map $\Gamma : A \to B$ is said to be upper hemi-continuous (uhc) at a point a if for any open neighborhood V of $\Gamma(a)$ there exists a neighborhood U of a such that $\Gamma(U) \subseteq V$.

If B is compact, uhc is equivalent to the graph of Γ being closed, meaning that if $a_n \to a^*$ and $b_n \in \Gamma(a_n) \to b^*$, then $b^* \in \Gamma(a^*)$.

Definition 2: The set valued map $\Gamma : A \to B$ is said to be lower hemi-continuous (lhs) at a point a if for any open neighborhood V intersecting $\Gamma(a)$ there exists a neighborhood U of a such that $\Gamma(x)$ intersects V for all $x \in U$.

Equivalently, for all $a_m \to a \in A$ and $b \in \Gamma(a)$, there exists a_{m_k} subsequence of a_m and a corresponding $b_k \in \Gamma(a_{m_k})$ such that $b_k \to b$.

A set-valued map is said to be continuous if it is both upper and lower hemi-continuous. A single-valued function is continuous if and only if it is uhc. We restate a version of the Berge Maximum Theorem with a compactness assumption from [13].

Lemma 2 (Berge Maximum Theorem): Let $\mathcal{A} \subseteq \mathbb{R}$ and $\mathcal{S} \subseteq \mathbb{R}^{m\times n}$. Assume that $J : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is jointly continuous and $\Gamma : \mathcal{A} \to \mathcal{S}$ is a compact-valued correspondence. Define

$$K^*(\alpha) = \arg\min\{J(K, \alpha) | K \in \Gamma(\alpha)\}, \text{ for all } \alpha \in \mathcal{A},$$

and

$$J(K^*(\alpha), \alpha) = \min\{J(K, \alpha) | K \in \Gamma(\alpha)\}, \text{ for all } \alpha \in \mathcal{A}.$$

If Γ is continuous at some $\alpha \in \mathcal{A}$, then $J(K^*(\alpha), \alpha)$ is continuous at α. Furthermore, K^* is non-empty, compact-valued, closed, and upper hemi-continuous.

The Berge Maximum Theorem does not trivially apply to ODC since the set of stabilizing controllers is open and often unbounded.

Theorem 1: Assume that $K^*(0)$ is non-empty. Then the set $K^*(\alpha)$ is non-empty for all $\alpha > 0$. Moreover, $K^*(\alpha)$ is upper hemi-continuous and the optimal cost $J(K^*(\alpha), \alpha)$ is continuous and strictly decreasing with respect to α.

Proof: When $K^*(0)$ is non-empty, there is an optimal decentralized controller for the undamped system. With the set of stabilizing controller non-empty, we invoke the "damping property" in Lemma 1 and conclude

$$J(K^*(\alpha), \alpha) \leq J(K^*(0), \alpha) < J(K^*(0), 0).$$

The inequality above assumed existence of a globally controller for all values of the damping parameter α. This is true because the lower-level set of $J(K, \alpha)$ is compact [17]. Precisely, define $\Gamma_M(\alpha)$ to be

$$\Gamma_M(\alpha) = \{K \in S : A - \alpha I + BK \text{ stable}, J(K, \alpha) \leq M\}. \tag{7}$$

The set-valued function Γ_M is compact-valued for all constant α given a fixed M. From the damping property, we can select any $M > J(K^*(0), 0)$ and optimize instead over $\Gamma_M(\alpha)$ without losing any globally optimal controller. The continuity of $\Gamma_M(\alpha)$ at α for almost all values of M is proved in the appendix. The Berge Maximum Theorem yields the desired continuity of $K^*(\alpha)$ and $J(K^*(\alpha), \alpha)$. ■

The argument above can be extended to characterize all locally optimal controllers. A caveat is the possible existence of locally optimal controllers with unbounded costs. Their existence does not contradict the damping property — damping can introduce locally optimal controllers that are not stabilizing without the damping.

Theorem 2: Assume that $K^\dagger(0)$ is non-empty. Then, the set $K^\dagger(\alpha)$ is nonempty for all $\alpha > 0$. Suppose furthermore that at an $\alpha_0 > 0$,

$$\lim_{\epsilon \to 0^+} \sup_{\alpha \in [\alpha_0 - \epsilon, \alpha_0 + \epsilon]} \sup_{K \in K^\dagger(\alpha)} J(K, \alpha) < \infty.$$

Then $K^\dagger(\alpha)$ is upper hemi-continuous at α_0 and the optimal cost $J(K^\dagger(\alpha), \alpha)$ is upper hemi-continuous at α_0.

Proof: The fact that $K^\dagger(\alpha)$ is non-empty follows from the existence of globally optimal controllers in Theorem 1. Consider the parametrized optimization problem

$$\begin{aligned} \min \quad & \|\nabla J(K, \alpha)\| \\ s.t. \quad & K \in \Gamma_M(\alpha), \end{aligned} \tag{8}$$

where $\|\cdot\|$ denotes the 2-norm of a vector. The assumption of the Lemma ensures the existence of an M and an $\epsilon > 0$ such that $M > J(K, \alpha)$ for $K \in K^\dagger(\alpha)$ where $\alpha \in [\alpha_0 - \epsilon, \alpha_0 + \epsilon]$. This choice of M guarantees that the formulation (8) does not cut off any locally optimal controllers. As proved in the appendix, $\Gamma_M(\alpha)$ is continuous at α_0 for almost all values of M, and a large M can be selected to make $\Gamma_M(\alpha)$ continuous at α_0. Berge Maximum Theorem applies to conclude that $K^\dagger(\alpha)$ is upper hemi-continuous. Since $J(K, \alpha)$ is jointly continuous in (K, α), $J(K^\dagger(\alpha), \alpha)$ is upper hemi-continuous. ■

IV. Asymptotic Properties

In this section, we prove asymptotic properties of the local solutions $K^\dagger(\alpha)$. The following theorem characterizes the evolution of locally optimal controllers for a specific sparsity pattern. The theorem justifies the practice of random initialization around zero.

Theorem 3: Suppose that the sparsity pattern I_S is block-diagonal with square blocks and that R has the same sparsity pattern as I_S. Then, all points in $K^\dagger$ converge to the zero matrix as $\alpha \to \infty$. Furthermore, $J(K, \alpha) \to 0$ as $\alpha \to \infty$ for all $K \in K^\dagger(\alpha)$.

Proof: Recall the expression of the objective function (2), the first-order necessary conditions (5a)-(5d), and (6). As α increases, some local solutions may disappear, some new local solutions may appear. The appearance cannot occur infinitely often because the equations (5a)-(5d) are algebraic. Suppose that the number of local solutions does not change when α is greater than some constant α_0. The damping property ensures for all $\beta > \alpha > \alpha_0$ that

$$\max_{K\in K^{\dagger}(\beta)} J(K,\beta) \le \max_{K\in K^{\dagger}(\alpha)} J(K,\beta).$$

The right-hand side optimizes over a fixed, finite set of controllers and approaches 0 as $\beta \to \infty$ due to the formulation (2) and the dominated convergence theorem. The left-hand side, therefore, also converges to zero as $\beta \to \infty$. From (6) and the assumption that D_0 is positive-definite, $\|P_\beta(K)\| \to 0$ for all $K \in K^{\dagger}(\beta)$ as $\beta \to \infty$.

The sparsity assumption allows the expression of the locally optimal controllers in (5c) as

$$K = -R^{-1}((B^\top P_\alpha(K) L_\alpha(K)) \circ I_S)(L_\alpha(K)\circ I_S)^{-1}.$$

In particular, we bound

$$\|BK\| \le \|BR^{-1}B^\top P_\alpha(K)L_\alpha(K)\| \lambda_{\min}(L_\alpha(K))^{-1}.$$

Pre- and post-multiplying (5b) by the unit eigenvector v of the smallest eigenvalue of $L_\alpha(K)$ yields

$$\lambda_{\min}(L_\alpha(K))(2a - 2v^\top(A+BK)v) = v^\top D_0 v. \tag{9}$$

Therefore,

$$\begin{aligned}\lambda_{\min}(L_\alpha(K)) &\ge \frac{\lambda_{\min}(D_0)}{2\alpha + 2\|A+BK\|}\\ &\ge \frac{\lambda_{\min}(D_0)}{2\alpha+2\|A\|+2\|BK\|}\\ &\ge \frac{\lambda_{\min}(D_0)}{2\alpha+2\|A\|+2\|BR^{-1}B^\top P_\alpha(K)L_\alpha(K)\|\lambda_{\min}(L_\alpha^{-1}(K))},\end{aligned}$$

which simplifies to

$$\lambda_{\min}(L_\alpha(K)) \ge \frac{\lambda_{\min}(D_0) - 2\|BR^{-1}B^\top P_\alpha(K)L_\alpha(K)\|}{(2\alpha+2\|A\|)}. \tag{10}$$

Take the trace of (5b) and consider the estimate

$$\begin{aligned}&2n\|A\|\|L_\alpha\| + \operatorname{tr}(D_0) \ge 2\|A\|\operatorname{tr}(L_\alpha) + \operatorname{tr}(D_0)\\ &\ge 2\alpha\operatorname{tr}(L_\alpha) + 2\operatorname{tr}(BR^{-1}((B^\top P_\alpha L_\alpha)\circ I_S)(L_\alpha\circ I_S)^{-1}L_\alpha)\\ &\ge 2\alpha\operatorname{tr}(L_\alpha) - 2\|BR^{-1}((B^\top P_\alpha L_\alpha)\circ I_S)\|\operatorname{tr}((L_\alpha\circ I_S)^{-1}L_\alpha)\\ &= 2\alpha\operatorname{tr}(L_\alpha) - 2\|BR^{-1}((B^\top P_\alpha L_\alpha)\circ I_S)\| n\\ &\ge 2\alpha\|L\| - 2n\|BR^{-1}\|\|B^\top\|\|P_\alpha\|\|L_\alpha\|,\end{aligned} \tag{11}$$

where for clarity we drop the implicit dependence on K in L_α and P_α. The second and the third inequalities use the fact that $|\operatorname{tr}(AL)| \le \|A\|\operatorname{tr}(L)$ for a positive-definite matrix L and an arbitrary matrix A. The estimate (11), combined with the previous argument that $\|P_\alpha\| \to 0$, concludes $\|L_\alpha\| \to 0$. We also obtain from the inequality (11) that

$$\|L_\alpha\| \le \frac{\operatorname{tr}(D_0)}{2a - 2n\|A\| - 2n\|BR^{-1}\|\|B^\top\|\|P_\alpha\|}, \tag{12}$$

for a small enough P_α. Combining (10) and (12) leads to

$$\begin{aligned}\|K\| &\le \|R^{-1}\|\cdot\|(B^\top P_\alpha L_\alpha)\circ I_S\|\cdot\|(L_\alpha\circ I_S)^{-1}\|\\ &\le \|R^{-1}\|\cdot\|B^\top\|\cdot\|P_\alpha\|\cdot\|L_\alpha\|\cdot\|\lambda_{\min}(P_\alpha)^{-1}\|\\ &\le \|R^{-1}\|\cdot\|B^\top\|\cdot\|P_\alpha\|\\ &\times\frac{\operatorname{tr}(D_0)}{2\alpha - 2n\|A\| - 2n\|BR^{-1}\|\|B^\top\|\|P_\alpha\|}\\ &\times\frac{(2\alpha+2\|A\|)}{\lambda_{\min}(D_0) - 2\|BR^{-1}B^\top P_\alpha L_\alpha\|},\end{aligned}$$

which converges to 0 as $\alpha \to \infty$. ■

Not only do all locally optimal controllers approach zero, the problem is also convex over bounded regions with enough damping.

Theorem 4: For any given $r > 0$, the Hessian matrix $\nabla^2 J(K,\alpha)$ is positive definite over $\|K\| \le r$ for all large α.

Proof: The proof requires the vectorized Hessian formula given in Lemma 3.7 of [15], restated below. We use $\otimes$ to denote the Kronecker project of two matrices and vec to denote the vectorized operation that stack the columns of a matrix together into a vector. Define $j_\alpha : \mathbb{R}^{m\cdot n} \to \mathbb{R}$ by $j_\alpha(vec(K)) = J(K,\alpha)$. The Hessian of j_α is given by the formula

$$H_\alpha(K) = 2\left\{(L_\alpha(K)\otimes R) + G_\alpha(K)^\top + G_\alpha(K)\right\}, \tag{13}$$

where

$$\begin{aligned}G_\alpha(K) =&[I\otimes(B^\top P_\alpha(K) + RK)]\times\\ &[I\otimes(A-\alpha I+BK) + (A-\alpha I+BK)\otimes I]^{-1}\\ &(I_{n,n} + P(n,n))[L_\alpha(K)\otimes B]\end{aligned}$$

and $P(n,n)$ is an $n^2\times n^2$ permutation matrix.

We first show that $H_\alpha(K)$ in (13) is positive definite for any fixed K when α is large. Recall the definition of L_α and P_α in (5a)-(5b) and apply the triangle inequality:

$$\begin{aligned}&2\alpha\|L_\alpha(K)\| \le \|D_0\| + 2\|A+BK\|\|L_\alpha(K)\|\\ &2\alpha\|P_\alpha(K)\| \le \|Q\| + 2\|A+BK\|\|P_\alpha(K)\| + \|R\|\|K\|^2,\end{aligned}$$

which implies that $\|P_\alpha(K)\| \to 0$ and $\|L_\alpha(K)\| \to 0$ as $\alpha\to\infty$. The minimum eigenvalue of $L_\alpha(K)$ can be bounded similarly: let v be the unit eigenvector of $L_\alpha(K)$ corresponding to $\lambda_{\min}(L_\alpha(K))$; pre- and post-multiplying (5b) by v, we obtain

$$\lambda_{\min}(L_\alpha(K)) \ge \frac{v^T D_0 v}{2\alpha - 2v^T(A+BK)v} \ge \frac{\lambda_{\min}(D_0)}{2\alpha+2\|A+BK\|}. \tag{14}$$

The first Hessian term $L_\alpha(K)\otimes R$ in (13) can bounded from below using (14).

$$\begin{aligned}\lambda_{\min}\left(L_\alpha(K)\otimes R\right) &= \lambda_{\min}(L_\alpha(K))\lambda_{\min}(R)\\ &\ge \frac{\lambda_{\min}(D_0)\lambda_{\min}(R)}{2\alpha+2\|A+BK\|}.\end{aligned}$$

We bound the norm of the second and the third Hessian term $\|G_\alpha(K)\|$ as follows:

$$\begin{aligned}\|G_\alpha(K)\| &\le \|I \otimes (B^T P_\alpha(K) + RK)\| \\ &\quad \times \| \left[I \otimes (A - \alpha I + BK) + (A - \alpha I + BK) \otimes I \right]^{-1} \| \\ &\quad \times \|(I_{n,n} + P(n,n))[L_\alpha(K) \otimes B]\| \\ &\lesssim (-\lambda_{\max} (I \otimes (A - \alpha I + BK) + (A - \alpha I + BK) \otimes I))^{-1} \\ &\quad \times \|L_\alpha(K)\| \\ &\lesssim (2\alpha)^{-1} \|L_\alpha(K)\|,\end{aligned}$$

where $\lesssim$ hides constants that do not depend on α. Comparing the two estimates above, we find that the first term $L_\alpha(K) \otimes R$ in (13) dominates its following terms $G_\alpha(K)^\top + G_\alpha(K)$ with a large α for all bounded K. Therefore, the Hessian $H_\alpha(K)$ is positive definite over bounded K when α is large. Note that $H_\alpha(K)$ is the Hessian of the objective function when the controller is centralized. The conclusion carries over the decentralized controller because the Hessian for the decentralized controller is a principal sub-matrix of the Hessian for the centralized controller. ■

Corollary 1: Under the assumption of Theorem 3 and Theorem 4, there is no spurious locally optimal controller for large α, or equivalently, $K^\dagger(\alpha) = K^*(\alpha)$ for all large values of α.

Proof: For any given $r > 0$, all controllers in the ball $\mathcal{B} = \{K : \|K\| \le r\}$ are stabilizing when α is large. As a result, stability constraints can be relaxed over $\mathcal{B}$. Furthermore, from Theorem 3, when α is large, all locally optimal controllers will be inside $\mathcal{B}$. From Theorem 4, the objective function becomes convex over $\mathcal{B}$ for large enough α. These observations imply that local and global solutions coincide. ■

Corollary 2: Under the same assumption of Theorem 3 and Theorem 4, suppose further that the globally optimal solution is unique for all damping parameters, namely, $K^*(\alpha)$ is a singleton set for all $\alpha \ge 0$. Then, the trajectory $K^*(\alpha)$ is continuous. Moreover, if there is an $\epsilon > 0$ such that the local search method initialized at ϵ distance away from $K^*(\alpha)$ converges to $K^*(\alpha)$, then Algorithm 1 and Algorithm 2 output the globally optimal stabilizing controller in $K^*(0)$ with a proper discretization of the α space.

Proof: We have shown in Theorem 1 that $K^*(\alpha)$ is upper hemi-continuous. With the singleton assumption, we conclude the continuity of $K^*(\alpha)$ because a single-valued function is continuous if and only if it is upper hemi-continuous. We choose a discretization $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_T$, where α_T is large enough for which the "no spurious property" of Corollary 1 holds. As a result, Algorithm 1 and Algorithm 2 are able to locate the continuous globally optimal trajectory $K^*(\alpha)$ at $\alpha = \alpha_T$. To obtain $K^*(0)$, we follow the continuous $K^*(\alpha)$ in the manner of Algorithm 1 and Algorithm 2, where α_t and α_{t+1} are close enough so K_{t+1} lies in the region where the local search method initialized at K_{t+1} converges to K_t. This discretization inductively yields a serious of controllers K_t, for $t = T, T-1, \ldots, 0$ that all lie on the path $K^*(\alpha)$, for $\alpha \in [0, \alpha_T]$. ■

Remark 1: A proper discretization $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_T$ has a large α_T for which the "no spurious property" of Corollary 1 holds. A proper discretization further requires α_t and α_{t+1} to be reasonably close to guarantee that the local search method initialized at K_{t+1} is able to converge to K_t in Algorithm 1 and Algorithm 2.

V. Numerical Experiments

This section documents various homotopy behaviors as the damping parameter α varies. The focus is on the evolution of locally optimal trajectories, which can be tracked by any local search method. The experiments are performed on small-sized systems so the random initialization can find a reasonable number of distinct locally optimal solutions. Despite the small system dimension, the existence of many locally optimal solutions and their convoluted trajectories demonstrate the efficiency of the proposed method for solving the optimal decentralized control problem.

The local search method used here is the projected gradient descent. At a controller K^i, we perform line search along the direction $\tilde{K}^i = -\nabla J(K) \circ I_S$. The step size is determined with backtracking and the Armijo rule, where we select s^i as the largest number in $\{\bar{s}, \bar{s}\beta, \bar{s}\beta^2, ...\}$ such that $K^i + s^i \tilde{K}^i$ is stabilizing while

$$J(K^i + s^i \tilde{K}^i) < J(K^i) + \gamma s^i \langle \nabla J(K^i), \tilde{K}^i \rangle.$$

Our choice of parameters is $\gamma = 0.001$, $\beta = 0.5$, and $\bar{s} = 1$. We terminate the iteration when the norm of the gradient is less than 10^{-3}.

A. Systems with a Large Number of Local Minima

We first consider an instance of the exponential class from [6], where the feasible set is highly disconnected and admits several local minima. The system matrices are given by

$$A = \begin{bmatrix} -1 & 2 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & -1 & 0 & 2 \\ 0 & 0 & -2 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \tag{15}$$

$$D_0 = I, \quad I_{\mathcal{S}} = I, \quad Q = I, \quad R = I. \tag{16}$$

When the dimension n is equal to 4, it is known that the set of stabilizing decentralized controllers has at least 5 connected components and hence at least 5 locally optimal controllers. We sample the initial controllers from the normal distribution with zero mean and unit variance and, after 1000 samples, obtain 5 initial locally optimal solutions. We gradually increase the damping parameter from 0 to 0.6 with a 0.002 increment, and track the trajectories of locally optimal solutions in the spirit of Algorithm 1. The evolution of the optimal cost and the distance from the best known optimal controller is plotted Figure 1. Notice that all sub-optimal local trajectories terminate after a modest damping $\alpha \approx 0.2$. After that, the minimization algorithm always tracks a single trajectory. This illustrates the prediction of Corollary 1. Especially, if we start tracking a sub-optimal controller trajectory from $\alpha = 0$, we will be on the better trajectory when $\alpha \approx 0.2$. At that time, if we gradually decrease α to zero, we obtain a stabilizing controller with a lower cost.

The above observation is valid for larger systems. An example is presented in the technical report [7], where among the 50 locally optimal trajectories Algorithm 1 successfully tracked the globally optimal trajectory.

B. Experiments on Random Systems

With the same initialization and optimization procedure, we perform the experiments on 3-by-3 system matrices A and B randomly generated from the normal distribution with zero mean and unit variance. For 92 out of 100 samples, we are not able to find more than one locally optimal trajectory. Examples with more than one local trajectories are listed below. The top plot in each figure shows the cost of locally optimal controllers. The bottom plot shows the distance of the locally optimal controllers to the controller with the lowest cost. Note that the order of the cost trajectories may be preserved during the damping (Figure 2), or may be disrupted (Figure 3). More than one trajectory may have the lowest cost as the damping increases (Figure 4), but with a high damp, there is only one trajectory that has the lowest cost. In Figure 3, at the intersection of the two curves, there are two distinct global solutions and therefore Algorithm 1 may fail to obtain the globally optimal decentralized controller. This illustrates the necessity of assuming the uniqueness of the globally optimal controller in Corollary 2.

VI. Conclusion

This paper showed that damping the system dynamics effectively reduces the number of locally optimal decentralized control policies. We proved the asymptotic and continuity properties of trajectories as the damping parameter varied. These property led to sufficient conditions under which the proposed local search methods were able to find the global solution of the optimal distributed control problem. The complicated phenomenon of continuation was illustrated with numerical examples.

Acknowledgments: The authors are grateful to Salar Fattahi and Cédric Josz for their constructive comments and feedback. The authors thank Yuhao Ding for sharing the implementation of local search algorithms.

References

[1] Vincent D. Blondel and John N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000.

[2] J. R. Broussard and N. Halyo. Active Flutter Control using Discrete Optimal Constrained Dynamic Compensators. In *1983 American Control Conference*, pages 1026–1034, June 1983.

[3] J.C. Doyle, K. Glover, P.P. Khargonekar, and B.A. Francis. State-space solutions to standard H-2 and H-infinity control problems. *IEEE Transactions on Automatic Control*, 34(8):831–847, 1989.

[4] Emmanuel G. Collins Jr. and Debashis Sadhukhan. A comparison of descent and continuation algorithms for H2 optimal, reduced-order control designs. *International Journal of Control*, 69(5):647–662, January 1998.

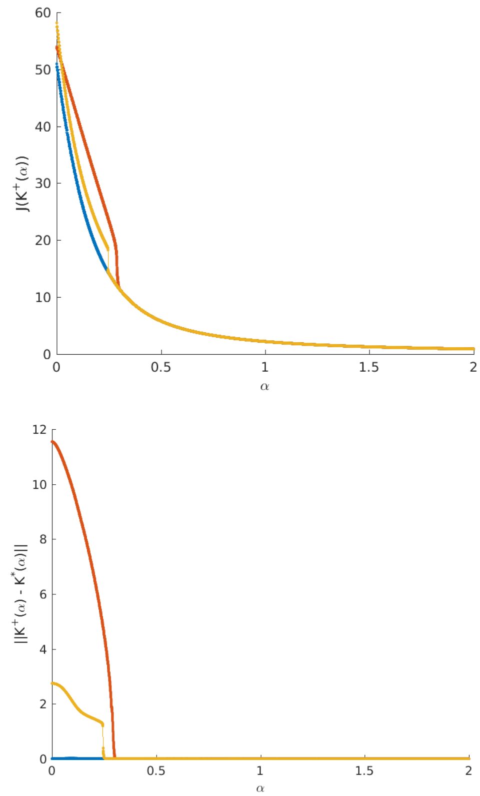


Fig. 2. Trajectories of a randomly generated system where the order of locally optimal controllers is preserved as the damping parameter α changes.

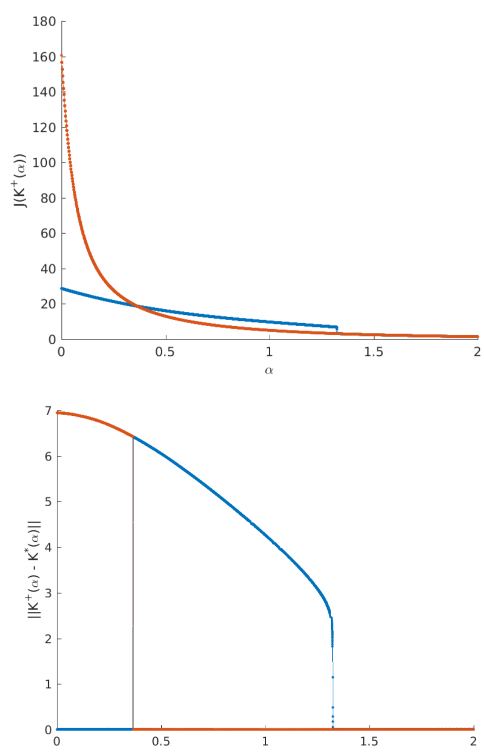


Fig. 3. Trajectories of a randomly generated system where the order of locally optimal controllers is disrupted as the damping parameter α changes.

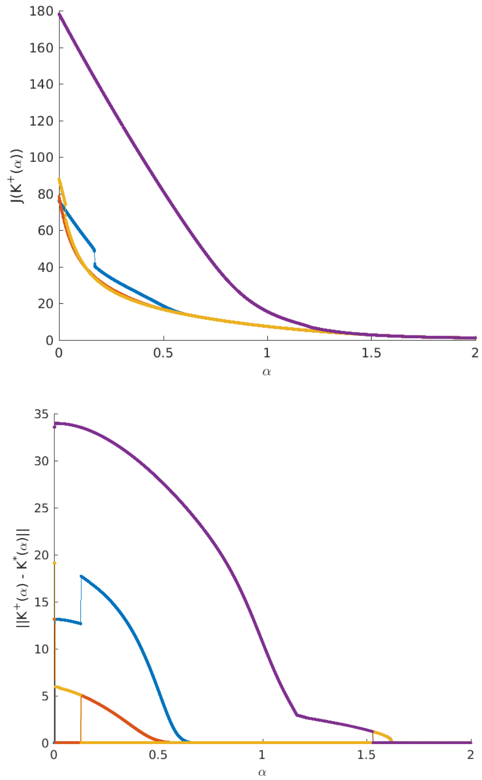


Fig. 4. Trajectories of a randomly generated system with a complicated behavior.

[5] Maryam Fazel, Rong Ge, Sham M. Kakade, and Mehran Mesbahi. Global Convergence of Policy Gradient Methods for Linearized Control Problems. In *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80, 2018.

[6] Han Feng and Javad Lavaei. On the Exponential Number of Connected Components for the Feasible Set of Optimal Decentralized Control Problems. In *Proceedings of the 2019 American Control Conference*, pages 1430–1437, Philadelphia, PA, USA.

[7] Han Feng and Javad Lavaei. Damping with Varying Regularization in Optimal Decentralized Control. https://lavaei.ieor.berkeley.edu/ODC_hom_2019_2.pdf, 2019.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[9] Moritz Hardt, Benjamin Recht, and Yoram Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, volume 48, pages 1225–1234, New York, NY, USA, September 2015.

[10] Laurent Lessard and Sanjay Lall. An algebraic approach to the control of decentralized systems. *IEEE Transactions on Control of Network Systems*, 1(4):308–317, 2014.

[11] Mathieu Mercadal. Homotopy approach to optimal, linear quadratic, fixed architecture compensation. *Journal of Guidance, Control, and Dynamics*, 14(6):1224–1233, 1991.

[12] Hossein Mobahi and John W. Fisher Iii. A Theoretical Analysis of Optimization by Gaussian Continuation. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, February 2015.

[13] Efe A. Ok. *Real Analysis with Economic Applications*. Princeton University Press, 2007.

[14] Anders Rantzer. Scalable control of positive systems. In *European Journal of Control*, volume 24, pages 72–80, 2015.

[15] T. Rautert and E. W. Sachs. Computational Design of Optimal Output Feedback Controllers. *SIAM Journal on Optimization*, 7(3):837–852, August 1997.

[16] Parikshit Shah and Pablo A. Parrilo. H2-optimal decentralized control over posets: A state-space solution for state-feedback. *IEEE Transactions on Automatic Control*, 58(12):3084–3096, 2013.

[17] H. T. Toivonen. A globally convergent algorithm for the optimal constant output feedback problem. *International Journal of Control*, 41(6):1589–1599, June 1985.

[18] Yuh Shyang Wang, Nikolai Matni, and John C. Doyle. System Level Parameterizations, constraints and synthesis. In *Proceedings of the American Control Conference*, pages 1308–1315, Seattle, WA, USA, 2017.

[19] H. S. Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6(1):131–147, February 1968.

[20] Dragan Zigic, Layne T. Watson, Emmanuel G. Collins, and Dennis S. Bernstein. Homotopy Approaches to the H_2 Reduced Order Model Problem. Departmental Technical Report TR 91-24, Department of Computer Science, Virginia Polytechnic Institute & State University, January 1991.

Appendix

In Lemma 3 and Lemma 4 below, we prove the continuity of the lower level-set map defined in (7).

Lemma 3: Assume that $\Gamma_M(\alpha)$ is not empty for all $\alpha \geq 0$ and a given $M > 0$, then $\Gamma_M(\alpha)$ is an upper hemi-continuous set-valued map.

Proof: From [17], $\Gamma_M(\alpha)$ is compact for all α. Due to the damping property, for any $\alpha < \beta$, we have $\Gamma_M(\alpha) \subseteq \Gamma_M(\beta)$. Therefore, to characterize the continuity of Γ at a point $\alpha^* \geq 0$, it suffices to consider the restricted map $\Gamma_M : [\alpha^* - \epsilon, \alpha^* + \epsilon] \to \Gamma_M(\alpha^* + \epsilon)$ for some $\epsilon > 0$, that is, to consider the range of Γ_M to be compact. Therefore, the sequence characterization of uhc applies. Suppose that $\alpha_i \to \alpha^*$, select a sequence of $K_i \in \Gamma_M(\alpha_i)$ that converges to K^*. The continuity of $J(K, \alpha)$ implies $J(K^*, \alpha^*) \leq M$. The fact that the cost is bounded implies that $A - \alpha^* I + BK$ is stable. Since the subspaces of matrices are closed, we have $K^* \in \mathcal{S}$. All conditions for $K^* \in \Gamma(\alpha^*)$ are verified, and therefore Γ_M is upper hemi-continuous. ∎

Lemma 4: At any given $\alpha^* \geq 0$, $\Gamma_M(\alpha)$ is lower hemi-continuous at α^* except when $M \in \{J(K, \alpha^*) : K \in K^\dagger(\alpha^*)\}$, which is a finite set of locally optimal costs.

Proof: Prove by contradiction. Consider a sequence $\alpha_i \to \alpha^*$ and a matrix $K^* \in \Gamma_M(\alpha^*)$, for which there exist no subsequence of α_i and $K_i \in \Gamma_M(\alpha_i)$ such that $K_i \to K^*$. We must have

- $J(K^*, \alpha^*) = M$ — otherwise by the continuity of J, $J(K^*, \alpha_i) < M$ for large i and, since the set of stabilizing controllers is open, $K^* \in \Gamma_M(\alpha_i)$ for large i, which is a contradiction.
- K^* must be a local minimum of $J(K, \alpha^*)$ — otherwise there exists a sequence $K_j \to K^*$ with $J(K_j, \alpha^*) < M$ and, by the continuity of J, there exists a sequence of large enough indices $n_j, j = 1, 2, \ldots$, such that $J(K_j, \alpha_{n_j}) < M$; the sequence $K_j \in \Gamma_M(\alpha_{n_j})$ converges to K^*.

The argument above implies that M is equal to the cost of some locally optimal controller at α^*. Because given α^*, $J(K, \alpha^*)$ can be described as a linear function in terms of K over an algebraic set given by (6), the cost of locally optimal controller take only finitely many values. ∎