Escaping Locally Optimal Decentralized Control Polices via Damping*

Han Feng and Javad Lavaei [†]

April 20, 2022

Abstract

We study the evolution of locally optimal decentralized controllers with the damping of the control system. Empirically it is shown that even for instances with an exponential number of connected components, damping merges all local solutions to the one global solution. We characterize the evolution of locally optimal solutions with the notion of hemi-continuity and further derive asymptotic properties of the objective function and of the locally optimal controllers as the damping becomes large. Especially, we prove that with enough damping, there is no spurious locally optimal controller with favorable control structures. The convoluted behavior of the locally optimal trajectory is illustrated with numerical examples.

1 Introduction

The optimal decentralized control problem (ODC) adds controller constraints to the classical centralized optimal control problem. This addition breaks down the separation principle and the classical solution formulas culminated in [4]. Although ODC has been proved intractable in general [23, 1], the problem has convex formulations under assumptions such as partially nestedness [19], positiveness [17], and quadratic invariance [10]. A recently proposed System Level Approach [21] convexified the problem in the space of system response matrix. Convex relaxation techniques have been extensively documented in [2], though it is considered challenging to solve large scale optimization problems with linear matrix inequalities.

The line of research on convexification is in contrast with the success of stochastic gradient descent well-documented in machine learning practice [9, 8]. Admittedly, the problem of generalizability, training speed, and fairness in machine learning departs from the traditional control focus on stability, robustness, and safety. Nevertheless, the interplay of the two has inspired fruitful results. As an example, to solve the linear-quadratic optimal control problem, the traditional nonlinear programming methods include Gauss-Newton, augmented

---

Lagrangian, and Newton's methods [11, 22, 12, 13]. Only in the last few years do researchers started to look at the classical problem with the newly developed optimization techniques and proved the efficiency of policy gradient methods in model-based and model-free optimal control problems [6]. This efficiency statement of local search, however, is unlikely to carry over trivially to ODC, due to the NP-hardness of the problem and the recent investigation of the topological properties of ODC in [7]. Nevertheless, questions can be answered without contradicting the general complexity statement. For example, it is known that damping of the system reduces the number of connected components of the set of stabilizing decentralized controllers. Does damping reduce the number of locally optimal decentralized controllers? This paper attempts an answer with (1) a study of the continuity properties of the trajectories of the locally optimal solutions formed by varying damping, and (2) an asymptotic analysis of the trajectories as the damping becomes large. The observation of our study shall shed light on the properties of local minima in reinforcement learning, whose aim is to design optimal control policies and different local minima have different practical behaviors.

This work is closely related to continuation methods such as homotopy. They are known to be appealing yet theoretically poorly understood [15]. Homotopy has been used as an initialization strategy in optimal control: in [3], the author mentioned the idea of gradually moving from a stable system to the original system to obtain a stabilizing controller. The paper [24] considered 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. [5] compared descent and continuation algorithms for 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 general constrained optimal control problem can be appreciated from the examples in [14]. Compared with those earlier works, we consider a special kind of continuation, that is, damping, to improve the locally optimal solutions in optimal decentralized control. Our focus is not so much on following a specific path but on the evolution of several paths and 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 2. Continuity and asymptotic properties of our damping strategies are outlined in Section 3 and Section 4, respectively. Numerical experiments are detailed in Section 5. Concluding remarks are drawn in Section 6.

2   Problem Formulation

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 $\mathbb{E}[x(0)x(0)^\top] = D_0$. 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 plane. The controller K is said to stabilize the system if $A + BK$ is stable. ODC optimizes over the set of structured stabilizing controllers

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

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 $K \circ I_\mathcal{S} = K$, where $\circ$ denotes entry-wise multiplication. In the following, we will consider the discounted, or damped cost, which is defined as

$$
\begin{aligned}
J(K, \alpha) =& \mathbb{E} \int_0^\infty \left[e^{-2\alpha t} \left(\hat{x}^\top(t) Q \hat{x}(t) + \hat{u}^\top(t) R \hat{u}(t) \right) \right] dt \\
s.t. \quad & \dot{\hat{x}}(t) = A\hat{x}(t) + B\hat{u}(t) \\
& \hat{u}(t) = K\hat{x}(t).
\end{aligned}
\tag{1}
$$

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

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

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 (1) and (2) satisfies the following "damping property": suppose that K stabilizes the system $(A - \alpha I, B)$, then for all $\beta > \alpha$, K stabilizes the system $(A - \beta I, B)$ with $J(K, \beta) < J(K, \alpha)$.*

3

Proof. From the formulation (3), 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. From formulation (1), $J(K, \beta) < J(K, \alpha)$. □

The ODC problem can be succinctly written as

$$
\begin{aligned}
\min \quad & J(K, \alpha) \\
s.t. \quad & K \in \mathcal{S} \\
& A - \alpha I + BK \text{ is stable.}
\end{aligned}
\tag{3}
$$

We denote its set of globally optimal controllers by $K^*(\alpha)$, and its set of locally optimal controllers by $K^\dagger(\alpha)$. The paper studies the properties of $K^*(\alpha)$, $K^\dagger(\alpha)$, and $J(K, \alpha)$ for $K \in K^*(\alpha)$ or $K^\dagger(\alpha)$.

To motivate the study of $K^\dagger(\alpha)$, consider Figure 1 below. The set-up of the experiments will be detailed in Section 5. It is known that systems of this type have a large number of locally optimal controllers [7]. The left figure 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 $J(K^*(\alpha), \alpha)$. The right figure plots the distance of the selected $K \in K^\dagger(\alpha)$ to the one $K \in K^*(\alpha)$.
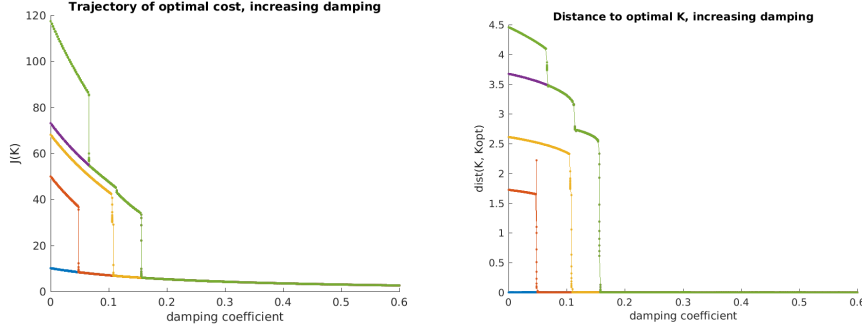

Figure 1: Trajectory of system in equation (11)

The fact that modest damping causes the locally optimal trajectories to "collapse" to each other is a very attractive phenomenon. Especially, they suggest two improving heuristics.

- Solve (3) from a large α and then gradually decrease α to 0.

- Start from a locally optimal $K \in K^\dagger(\alpha)$, solve (3) while gradually increase α to a positive value and then decrease α to 0.

The first idea shall avoid many unnecessary local optimum and its empirical behavior has been documented in [24]. The second idea has the potential to improve the locally optimal controllers obtained from many other methods. Due to the NP-hardness of general ODC, we expect no guarantee of producing a

globally optimal, or even a stabilizing, decentralized controller. The breakdown of these heuristics will be discussed in Section 5.

3 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.

Definition 1. *The set valued map $\Gamma : \mathcal{A} \to \mathcal{B}$ 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 $\Gamma(U) \subseteq V$.*

A related notion of lower hemi-continuity is provided in the supplement. 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 Berge Maximum Theorem with a compactness assumption from [16].

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*

$$K^*(\alpha) = \arg\min\{J(K,\alpha)|K \in \Gamma(\alpha)\}, \text{ for all } \alpha \in \mathcal{A}, \tag{4}$$

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.

Berge Maximum Theorem does not trivially apply to ODC: the set of stabilizing controllers is open and often unbounded. However, a lower-level set trick applies.

Theorem 1. *Assume that $K^*(0)$ is non-empty, then the set $K^*(\alpha)$ is non-empty for all $\alpha > 0$. $K^*(\alpha)$ is upper hemi-continuous and the optimal cost $J(K^*(\alpha), \alpha)$ is continuous and strictly decreasing in 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 incur 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 the globally controller for all values of damping parameter α. This is true because the lower-level set of $J(K, \alpha)$ is compact [20]. Precisely, define $\Gamma_M(\alpha)$ to be

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

5

The set-valued function Γ_M is compact-valued for all fixed 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 M is proved in the supplement. Berge maximum theorem then applies and yields the desired continuity of $K^*(\alpha)$ and $J(K^*(\alpha), \alpha)$. $\qquad\square$

The argument above can be extended to characterize all locally optimal controllers. A caveat is the possible existence of locally optimal controllers with unbounded cost. 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. That $K^\dagger(\alpha)$ is non-empty follows from the existence of globally optimal controllers in Theorem 1. Consider the parametrized optimization problem

$$\begin{aligned}
\min \quad & \|\nabla J(K, \alpha)\| \\
s.t. \quad & K \in \Gamma_M(\alpha).
\end{aligned} \tag{5}$$

The assumption 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 (5) does not cut off any locally optimal controllers. As proved in the supplement, $\Gamma_M(\alpha)$ is continuous at α_0 for almost any 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. $\qquad\square$

4 Asymptotic Properties

In this section, we state asymptotic properties of the local solutions $K^\dagger(\alpha)$. The controllers $K \in K^\dagger(\alpha)$ satisfy the first order necessary conditions in the following equations (6)-(9); their derivation can be found in [18].

$$(A - \alpha I + BK)^\top P_\alpha(K) + P_\alpha(K)(A - \alpha I + BK) + K^\top RK + Q = 0 \tag{6}$$

$$L_\alpha(K)(A - \alpha I + BK)^\top + (A - \alpha I + BK)L_\alpha(K) + D_0 = 0 \tag{7}$$

$$((B^\top P_\alpha(K) + RK)L_\alpha(K)) \circ I_S = 0 \tag{8}$$

$$K \circ I_S = K. \tag{9}$$

The above conditions provide a closed-form expression of the cost

$$J(K, \alpha) = \text{tr}(D_0 P_\alpha(K)). \tag{10}$$

It is worth pointing out that equations (6)-(10) 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 (6) and (7) given a stabilizing controller K. 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 $K \in K^\dagger(\alpha)$.*

Not only do all locally optimal controllers approach zero, the problem is in fact 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 $\|K\| \leq r$ for all large α.*

The proof of the two theorems above is given in the supplement.

Corollary 1. *With the assumption of Theorem 3, there is no spurious locally optimal controller for large α. That is, $K^\dagger(\alpha) = K^*(\alpha)$ for all large α.*

Proof. For any given $r > 0$, all controllers in the ball $\mathcal{B} = \{K : \|K\| \leq r\}$ 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 become convex over $\mathcal{B}$ for large enough α. The observations imply local and global solutions coincide. $\square$

The theorems above rely on the "damping property" in Lemma 1. It is worth commenting that damping the system with $-I$ is almost the only continuation method for general system matrices A that achieves the monotonic increasing of stable sets. Formally,

Theorem 5. *When $n \geq 3$, for any n-by-n real matrix H that is not a multiple of $-I$, there exists a stable matrix A for which $A + H$ is unstable.*

The proof is given in the supplement. This theorem justifies the use of $-\alpha I$ as the continuation parameter. However, in a given system with structure, matrices other than $-I$ may be appropriate.

5   Numerical Experiments

In this section, we document 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 methods. 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 demonstrates what is possible in a theory of homotopy.

The local search methods we used is the simplest 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 Armijo rule, that is, 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
$$J(K^i + s^i \tilde{K}^i) < J(K^i) + \alpha s^i \langle \nabla J(K^i), \tilde{K}^i \rangle.$$
Our choice of parameters are $\alpha = 0.001$, $\beta = 0.5$, and $\bar{s} = 1$. We terminate the iteration when the norm of the gradient is less than 10^{-3}.

5.1   Systems with a large number of local minima

We first consider the examples from [7], where the feasible set is reasonably disconnected and admits many local minima. The system matrices are given by

$$A = \begin{bmatrix} -1 & 2 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & -1 & 0 & 2 \\ 0 & 0 & -2 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, D_0 = I, \ I_S = I. \quad (11)$$

When the dimension n is 4, it is known that the set of stabilizing decentralized controllers has at least 5 connected components. We sample the initial controllers from $N(0, 1)$ and, after 1000 samples, obtain 5 initial optimal solutions. We gradually increase the damping parameter from 0 to 0.6 with 0.002 increment, and track the trajectories of locally optimal solutions by solving the newly damped system with the previous local optimal solution as the initialization. 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.

5.2   Experiments on Random Systems

With the same initialization and optimization procedure, we perform the experiments with 3-by-3 system matrices A and B randomly generated from the

distribution $N(0, 1)$. 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. All figures to the left plot the cost of locally optimal controllers. All figures to the right plot the distance of the locally optimal controllers to the controller with the lowest cost. Note that the order of the cost of the trajectories may be preserved during the damping (Figure 2) and may also be disrupted (Figure 3). More than one trajectory may have the lowest cost during the damping (Figure 4).
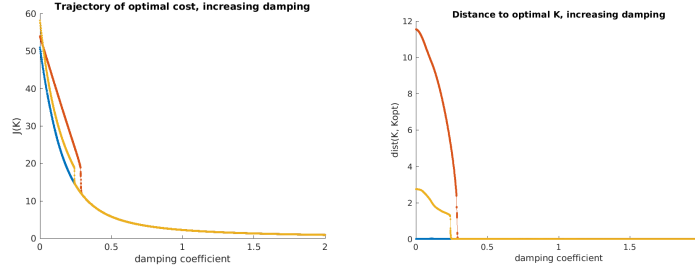

Figure 2: Trajectory of a randomly generated system where the order of locally optimal controller is preserved.
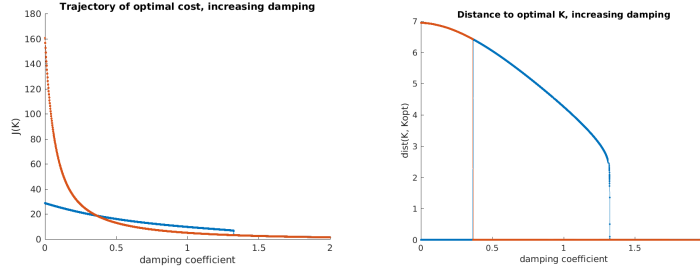

Figure 3: Trajectories of a randomly generated system where the order of locally optimal controller is disrupted.

Figure 5 shows a hysteresis-like loop as the damping coefficient is first decreased and then increased. The trajectory of the controller first leads up to large cost and, the local search method escapes this local minimum to another one with a smaller cost. As the damping decreases, it returns where it starts along a different route.

6   Conclusion

This paper studied the trajectory of locally and globally optimal solution to the optimal decentralized control problem as the damping of the decentralized
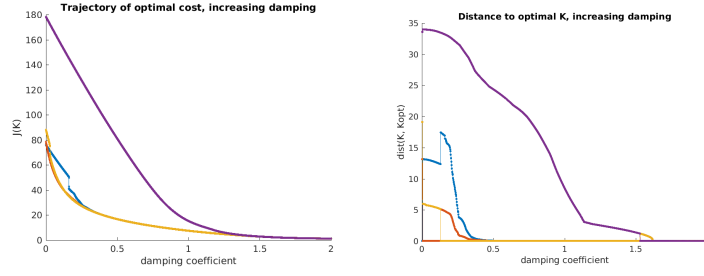
Figure 4: Trajectory of a randomly generated system with a complicated behavior.
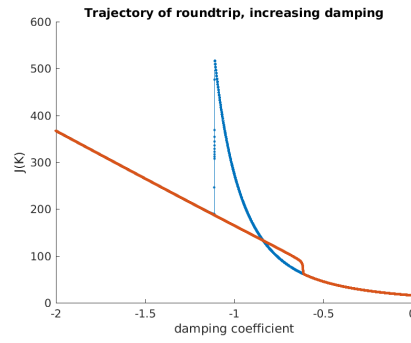

Figure 5: First decrease damping and then increase damping.

control system varies. Asymptotic and continuity properties of trajectories are proved. The complicated phenomenon of continuation is illustrated with numerical examples. The fact that damping merges all locally optimal solutions is strong evidence that the idea of homotopy can be fruitfully used to improve locally optimal solutions.

Acknowledgments

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] Stephen P. Boyd, L El Ghaoui, E Feron, and V Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15. 1994.

[3] 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.

[4] 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.

[5] 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.

[6] Maryam Fazel, Rong Ge, Sham M. Kakade, and Mehran Mesbahi. Global Convergence of Policy Gradient Methods for Linearized Control Problems. January 2018.

[7] Han Feng and Javad Lavaei. On the Exponential Number of Connected Components for the Feasible Set of Optimal Decentralized Control Problems. In *To Appear in Proceedings of the 2019 American Control Conference*, page 8.

[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. *arXiv:1509.01240 [cs, math, stat]*, 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] W. Levine and M. Athans. On the determination of the optimal constant output feedback gains for linear multivariable systems. *IEEE Transactions on Automatic Control*, 15(1):44–48, February 1970.

[12] Fu Lin, Makan Fardad, and Mihailo R. Jovanovic. Augmented Lagrangian Approach to Design of Structured Optimal State Feedback Gains. *IEEE Transactions on Automatic Control*, 56(12):2923–2929, December 2011.

[13] P. Makila and H. Toivonen. Computational methods for parametric LQ problems–A survey. *IEEE Transactions on Automatic Control*, 32(8):658–671, August 1987.

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

[15] 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.

[16] Efe A. Ok. *Real Analysis with Economic Applications*, volume 10. Princeton University Press, 2007.

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

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

[19] Parikshit Shah and Pablo A. Parrilo. $\mathscr{H}$2-optimal decentralized control over posets: A state-space solution for state-feedback. *IEEE Transactions on Automatic Control*, 58(12):3084–3096, 2013.

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

[21] 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, 2017.

[22] C. Wenk and C. Knapp. Parameter optimization in linear systems with arbitrarily constrained controller structure. *IEEE Transactions on Automatic Control*, 25(3):496–500, June 1980.

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

[24] Dragan Zigic, Layne T. Watson, Emmanuel G. Collins, and Dennis S. Bernstein. Homotopy Approaches to the H2 Reduced Order Model Problem. http://eprints.cs.vt.edu/archive/00000269/, January 1991.

A   Notions of continuity

We recount the notion of upper and lower hemi-continuity and prove the continuity properties of the lower level-set map. The reader is referred to [16] for an accessible treatment.

Definition 2. *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, that is, if $a_n \to a^*$ and $b_n \in \Gamma(a_n) \to b^*$, then $b^* \in \Gamma(a^*)$.

Definition 3. *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 $b_k \to b$.

We prove the upper hemi-continuity of the lower level set map in Lemma 3 below.

Lemma 3. *Given matrices A, B and the objective cost $J(K, \alpha)$ that satisfies the damping property. Define*

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

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

Proof. From [20], $\Gamma_M(\alpha)$ is compact for all α. From 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 $\alpha^* \ge 0$, 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 $\alpha_i \to \alpha^*$, pick a sequence of $K_i \in \Gamma_M(\alpha_i)$ that converges to K^*. The continuity of $J(K, \alpha)$ implies $J(K^*, \alpha^*) \le M$. The fact that the cost is bounded implies $A - \alpha^* I + BK$ is stable. Since subspaces of matrices are closed, $K^* \in \mathcal{S}$. We have verified all conditions for $K^* \in \Gamma(\alpha^*)$, so Γ_M is upper hemi-continuous. □

The lower hemi-continuity of Γ_M is more subtle.

Lemma 4. *At any given $\alpha^* \ge 0$, $\Gamma_M(\alpha)$ 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 $K^* \in \Gamma(\alpha^*)$, but there exists no subsequence of α_i and $K_i \in \Gamma(\alpha_i)$ such that $K_i \to K^*$. We must have $J(K^*, \alpha^*) = M$ — otherwise $J(K^*, \alpha_i) < M$ for large i and, since the set of stabilizing controllers is open, $K^* \in \Gamma_M(\alpha_i)$ for large i. Furthermore, 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 as sequence of large enough indices n_j 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 suggests that M belongs to the cost locally optimal controllers at α^*. Because $J(K, \alpha^*)$ as a function over K can be described as a linear function over an algebraic set, the value of local minimum is finite. □

B Convergence of locally optimal controllers

We prove the asymptotic properties of the locally optimal controllers in Section 4 of the main paper.

Theorem. *Suppose the sparsity pattern I_S is block-diagonal with square blocks, and R has the same sparsity pattern as I_S. Then all points in $K^\dagger(\alpha)$ converges 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

$$J(K, \alpha) = \mathbb{E} \int_0^\infty \left[e^{-2\alpha t} \left(\hat{x}^\top(t) Q \hat{x}(t) + \hat{u}^\top(t) R \hat{u}(t) \right) \right] dt$$

$$s.t. \quad \dot{\hat{x}}(t) = A\hat{x}(t) + B\hat{u}(t)$$

$$\hat{u}(t) = K\hat{x}(t), \tag{12}$$

and the first order necessary conditions

$$(A - \alpha I + BK)^\top P_\alpha(K) + P_\alpha(K)(A - \alpha I + BK) + K^\top RK + Q = 0 \tag{13}$$

$$L_\alpha(K)(A - \alpha I + BK)^\top + (A - \alpha I + BK)L_\alpha(K) + D_0 = 0 \tag{14}$$

$$((B^\top P_\alpha(K) + RK)L_\alpha(K)) \circ I_S = 0 \tag{15}$$

$$K \circ I_S = K. \tag{16}$$

Those first order conditions can be used to characterize the objective function

$$J(K, \alpha) = \operatorname{tr}(D_0 P_\alpha(K)). \tag{17}$$

As α increases, some local solution may disappear, some new local solution may appear. The appearance cannot happen infinitely often because the equations (13)-(16) are algebraic. Suppose when $\alpha \geq \alpha_0$, the number of local solutions does not change. The damping property ensures for $\beta > \alpha > \alpha_0$,

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

The right hand side optimizes over a fixed, finite set of controllers and goes to zero as $\beta \to \infty$ from the formulation (12) and the dominated convergence theorem. The left hand side, therefore, also converges to zero as $\beta \to \infty$. From (17) 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 assumption on sparsity allows the expression of the locally optimal controllers in (15) as

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

Especially we can bound

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

Pre- and post- multiply (14) by $L_\alpha(K)$'s unit minimum eigenvector v,

$$\lambda_{\min}(L_\alpha(K))(2a - 2v^\top(A + BK)v) = v^\top D_0 v. \tag{18}$$

Therefore

$$\lambda_{\min}(L_\alpha(K)) \geq \frac{\lambda_{\min}(D_0)}{2\alpha + 2\|A + BK\|} \tag{19}$$

$$\geq \frac{\lambda_{\min}(D_0)}{2\alpha + 2\|A\| + 2\|BK\|} \tag{20}$$

$$\geq \frac{\lambda_{\min}(D_0)}{2\alpha + 2\|A\| + 2\|BR^{-1}B^\top P_\alpha(K) L_\alpha(K)\| \lambda_{\min}(L_\alpha(K))^{-1}}. \tag{21}$$

This simplifies to

$$\lambda_{\min}(L_\alpha(K)) \geq \frac{\lambda_{\min}(D_0) - 2\|BR^{-1}B^\top P_\alpha(K) L_\alpha(K)\|}{(2\alpha + 2\|A\|)} \tag{22}$$

14

Take the trace of (14) and consider the estimate

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

where for clarity L_α denotes $L_\alpha(K)$ and P_α denotes $P_\alpha(K)$. The second and the third inequalities use the fact that $|\mathrm{tr}(AL)| \leq \|A\|\,\mathrm{tr}(L)$ for a positive definite matrix L and any matrix A. This estimate, combined with previous argument that $\|P_\alpha\| \to 0$, concludes $\|L_\alpha\| \to 0$. We also obtain from the inequality that

$$
\|L_\alpha\| \leq \frac{\mathrm{tr}(D_0)}{2a - 2n\|A\| - 2n\|BR^{-1}\|\|B^\top\|\|\|P_\alpha\|}, \tag{23}
$$

for small enough P_α. Combining (22) and (23)

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

which converges to 0 as $\alpha \to \infty$. $\qquad\square$

C   The Positive Definiteness of Hessian

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

Proof. The proof requires the vectorized Hessian formula given in Lemma 3.7 of [18], restated below.

Lemma 5 ([18]). *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*

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

where

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

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

We show that $H_\alpha(K)$ in the lemma is positive definite for any fixed K when α is large. Recall the definition of L_α and K_α.

$$
L_\alpha(K)(A - \alpha I + BK)^T + (A - \alpha I + BK)L_\alpha(K) + D_0 = 0, \tag{24}
$$

$$
P_\alpha(K)(A - \alpha I + BK) + (A - \alpha I + BK)^T P_\alpha(K) + K^T RK + Q = 0. \tag{25}
$$

15

With triangle inequality

$$2\alpha \|L_\alpha(K)\| \leq \|D_0\| + 2\|A + BK\| \|L_\alpha(K)\|$$
$$2\alpha \|P_\alpha(K)\| \leq \|Q\| + 2\|A + BK\| \|P_\alpha(K)\| + \|R\| \|K\|^2$$

which means $\|P_\alpha(K)\| \to 0$ 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- multiply (24) by v, we obtain

$$\lambda_{\min}(L_\alpha(K)) \geq \frac{v^T D_0 v}{2\alpha - 2v^T(A + BK)v} \geq \frac{\lambda_{\min}(D_0)}{2\alpha + 2\|A + BK\|}. \tag{26}$$

The first Hessian term $L_\alpha(K) \otimes R$ can bounded from below with (26)

$$\lambda_{\min}(L_\alpha(K) \otimes R) = \lambda_{\min}(L_\alpha(K))\lambda_{\min}(R) \geq \frac{\lambda_{\min}(D_0)\lambda_{\min}(R)}{2\alpha + 2\|A + BK\|}.$$

We bound the norm of the second and the third Hessian term $\|G_\alpha(K)\|$ as follows, where $\lesssim$ hides constants that do not depend on α.

$$\|G_\alpha(K)\| \leq \|I \otimes (B^T P_\alpha(K) + RK)\|$$
$$\cdot \| [I \otimes (A - \alpha I + BK) + (A - \alpha I + BK) \otimes I]^{-1} \| \cdot \|(I_{n,n} + P(n,n))[L_\alpha(K) \otimes B]\|$$
$$\lesssim (-\lambda_{\max}(I \otimes (A - \alpha I + BK) + (A - \alpha I + BK) \otimes I))^{-1} \|L_\alpha(K)\|$$
$$\lesssim (2\alpha)^{-1} \|L_\alpha(K)\|.$$

Comparing the two estimates above, we find the first term dominates the two following terms with large α, uniformly over bounded K. The Hessian $H_\alpha(K)$ is therefore positive definite over bounded K when α is large. The conclusion carries over to the Hessian of the decentralized controller, which is a principal sub-matrix of the Hessian of the centralized controller. $\qquad \square$

D   The uniqueness of the continuation direction

This section aims to prove the following result

Theorem. *When $n \geq 3$, for any n-by-n real matrix H that is not a multiple of $-I$, there exists a stable matrix A for which $A + H$ is unstable.*

Define the set of stable directions

$$\mathcal{H} = \{H : A + tH \text{ is stable whenever } A \text{ is stable and } t \geq 0\}, \tag{27}$$

where A and H are n-by-n real matrices.

Lemma 6. *All matrices in $\mathcal{H}$ is similar to a diagonal matrix with non-positive diagonal entries. Especially, they cannot have complex eigenvalues.*

Proof. When t is large, $A + tH$ is a small perturbation of tH, hence the eigenvalues of H has to be in the closed left half plane. With a suitable similar transform assume H is in real Jordan form. First consider the case of two by two matrices, and we denote the matrices by H_2 and A_2. Assume for contradiction that H_2 is not diagonalizable. The non-diagonal real Jordan form of H_2 has the following possibilities:

- $H_2 = \begin{bmatrix} h & 1 \\ 0 & h \end{bmatrix}$, where H_2 has real eigenvalues $h < 0$. Pick $A_2 = \begin{bmatrix} 4h & -2 \\ 10h^2 & -3h \end{bmatrix}$, which is stable because $tr(A_2) = h < 0$ and $\det(A_2) = 8h^2 > 0$. We have $A_2 + tH_2 = \begin{bmatrix} ht + 4hby & t-2 \\ 10h^2 & ht - 3h \end{bmatrix}$, whose stability criterion $tr(A_2 + tH_2) < 0$ and $\det(A_2 + tH_2) > 0$ amounts to

$$2ht + h < 0$$
$$h^2(t^2 - 9t + 8) > 0,$$

  or equivalently $t \in (-1/2, 1) \cup (8, +\infty)$. Especially when $t = 2$, $A_2 + tH_2$ is not stable.

- $H_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Pick a stable matrix $A = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}$. $A + tH$ is not stable when $t = 2$.

- $H_2 = \begin{bmatrix} 0 & f \\ -f & 0 \end{bmatrix}$, where $f > 0$, Pick $A = \begin{bmatrix} -1 & -4 \\ 1 & -1 \end{bmatrix}$, $A + \frac{2}{f}H_2 = \begin{bmatrix} -1 & -2 \\ -1 & -1 \end{bmatrix}$ is not stable.

- $H_2 = \begin{bmatrix} h & f \\ -f & h \end{bmatrix}$, where $h < 0$ and $f > 0$. By rescaling assume $f = 1$. Consider the following matrix function

$$G(t) = \begin{bmatrix} 0 & \frac{1}{2} + (u+w)h \\ -\frac{1}{2} + (u-w)h & h \end{bmatrix} + t \begin{bmatrix} h & 1 \\ -1 & h \end{bmatrix} \tag{28}$$

  We have

$$tr(G(t)) = h + 2ht$$
$$\det(G(t)) = (1 + h^2)t^2 + (1 + h^2 + 2hw)t + h^2(w^2 - u^2) + hw + \frac{1}{4}.$$

  Espeically,

$$tr(G(-\frac{1}{2})) = 0$$
$$\frac{d}{dt}trG(t) = 2h$$
$$\det(G(-\frac{1}{2})) = h^2(-\frac{1}{4} - u^2 + w^2)$$
$$\frac{d}{dt}\det G(t)\Big|_{t=-\frac{1}{2}} = 2hw$$

  Hence as long as

$$w > 0 \text{ and } -\frac{1}{4} - u^2 + w^2 > 0 \tag{29}$$

  for small enough $\epsilon > 0$, $A_2 = G(-\frac{1}{2} + \epsilon)$ is a stable matrix and there will be matrices $G(t)$ with $t > -\frac{1}{2}$ whose trace is negative and whose determinant is smaller. Consider the minimal value the determinant can take

$$\det G\left(-\frac{1}{2} - \frac{hw}{1+h^2}\right) = h^2\left(-\frac{1}{4} - u^2 + \frac{h^2}{1+h^2}w^2\right)$$

17

which means when

$$-\frac{1}{4} - u^2 + \frac{h^2}{1+h^2} w^2 < 0 \tag{30}$$

The matrix $G(t)$ with $t = -\frac{1}{2} - \frac{hw}{1+h^2}$ is unstable. There certainly exist u and w that satisfies (29) and (30).

For general n, H's real Jordan form is an block upper-triangular matrix

$$H = \begin{bmatrix} H_2 & * \\ 0 & * \end{bmatrix}$$

where H_2 can take the four possibilities mentioned above. We take the corresponding stable A_2 constructed above, which has the property that $A_2 + t_0 H_2$ is not stable for some $t_0 > 0$. Form the block diagonal matrix

$$A = \begin{bmatrix} A_2 & 0 \\ 0 & -I \end{bmatrix}$$

Then A is stable, while $A + t_0 H = \begin{bmatrix} A_2 + t_0 H_2 & * \\ 0 & * \end{bmatrix}$ is not stable. $\qquad\square$

We can strengthen the argument above and further characterize $\mathcal{H}$ in the case $n \geq 3$.

Lemma 7. *When $n \geq 3$, the set of stable directions $\mathcal{H}$ does not contain any matrices of rank 1, 2, ..., $n-2$.*

Proof. From lemma 6, we only need to consider the case where H is diagonal with negative diagonal entries. Assume there is a rank one matrix $H \in \mathcal{H}$, write

$$H = \begin{bmatrix} H_3 & 0 \\ 0 & * \end{bmatrix},$$

where $H_3 = \mathrm{diag}(-1, 0, 0)$. This is possible with the rank assumption. We will construct a stable 3-by-3 matrix A_3, such that there is some $t_0 > 0$ that makes $A_3 + t_0 H_3$ unstable, and then carry the instability to $A + t_0 H$ with the extended matrix

$$A = \begin{bmatrix} A_3 & 0 \\ 0 & -I \end{bmatrix}.$$

From [7], the set

$$T = \left\{ t : \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 5 & 1 & -1 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 0.85 & 0.2 & 0.2 \end{bmatrix} \text{ is stable} \right\}$$

has two disconnected components. Consider the Jordan decomposition of the matrix

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 0.85 & 0.2 & 0.2 \end{bmatrix} = P \, \mathrm{diag}(-0.2, 0, 0) P^{-1},$$

18

where P is some invertible matrix. Write

$$G(t) = 5P^{-1} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 5 & 1 & -1 \end{bmatrix} P + t \times \operatorname{diag}(-1, 0, 0).$$

After this similar transform, the set T can be written with $G(t, 0)$.

$$T = \{t : G(t) \text{ is stable}\}$$

Since T is disconnected there exists some $t_1 < t_2$ such that $G(t_1)$ is stable, while $G(t_2)$ is unstable with some eigenvalue in the right half plane. Setting $A_3 = G(t_1)$ and $t_0 = t_2 - t_1$ completes the proof. $\qquad\square$

Since we can perturb the direction and make H full-rank, the fact that H has rank one is not the substantial property. This is indeed the case.

Lemma 8. *When $n \geq 3$, $\mathcal{H} = \{-\lambda I, \lambda \geq 0\}$.*

Proof. From lemma 6, we only need to consider the case where H is diagonal with negative diagonal entries. Write

$$H = \begin{bmatrix} H_3 & 0 \\ 0 & * \end{bmatrix},$$

where $H_3 = \operatorname{diag}(h_1, h_2, h_3)$. The diagonal entries $h_i, i = 1, 2, 3$ are non-positive and not all equal. We will construct an A_3 and a corresponding t_0 such that A_3 is stable while $A_3 + t_0 H_3$ is not stable, and extend to the general A as in Lemma 7. The case where H_3 has rank 1 has been considered in Lemma 7. We show the remaining rank is impossible. Without loss of generality we rescale H_3 and assume $h_1 = -1$.

- $H_3 = \operatorname{diag}(-1, h_2, 0)$, where $h_2 < 0$. Consider the matrix function

$$G(t) = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -h_2 \\ 2 & 1 & 0 \end{bmatrix} + tH_3 = \begin{bmatrix} -t & -1 & 0 \\ 0 & th_2 & -h_2 \\ 2 & 1 & 0 \end{bmatrix}.$$

The characteristic polynomial of $G(t)$ is

$$\phi_{G(t)}(x) = x^3 + (t - th_2)x^2 + (h_2 - t^2 h_2)x + (t - 2)h_2.$$

The Routh-Hurwitz Criterion insists

$$t(1 - h_2) > 0$$
$$(t - 2)h_2 > 0$$
$$t(1 - h_2)h_2(1 - t^2) > (t - 2)h_2.$$

which is simplified with $h_2 < 0$ to

$$0 < t < 2 \tag{31}$$
$$(1 - h_2)t^3 + th_2 - 2 > 0. \tag{32}$$

Especially, when $t = \frac{3}{2}$, (32) simplifies to the obvious expression $\frac{1}{8}(11 - 15h_2) > 0$. when $t = 3$, (31) implies $G(t)$ is not stable. Setting $A_3 = G(\frac{3}{2})$ and $t_0 = \frac{3}{2}$ concludes the proof.

- $H_3 = \mathrm{diag}(-1, h_2, h_3)$, where without loss of generally we assume

$$-1 \le h_2, h_3 < 0, \text{ and one of them is not } -1. \tag{33}$$

Consider the matrix

$$G(t) = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & h_2 \\ ah_3 & h_3 & 0 \end{bmatrix} + tH_3 = \begin{bmatrix} -t & -1 & 0 \\ 0 & th_2 & h_2 \\ ah_3 & h_3 & th_3 \end{bmatrix}$$

Its Routh-Hurwitz Criterion insists

$$t > 0$$

$$f_1(t) = a - t + t^3 > 0 \tag{34}$$

$$f_2(t) = -ah_2h_3 + th_2h_3(h_2 + h_3) + t^3(1 - h_2)(1 - h_3)(-h_2 - h_3) > 0 \tag{35}$$

We claim that when

$$\sqrt{\frac{h_2h_3(h_2 + h_3)^2}{(-h_2 - h_3 + h_2h_3)^3}} < a < \sqrt{\frac{4}{27}} \tag{36}$$

the set of t that satisfy Routh-Hurwitz Criterion is disconnected. To see this, write the positive local minimum of $f_1(t)$ in (34) as $t_1 = \sqrt{\frac{1}{3}}$, and write the positive local minimum of $f_2(t)$ in (35) as $t_2 = \sqrt{\frac{h_2h_3}{3(1-h_1)(1-h_2)}}$. The condition (33) ensures that $t_1 < t_2$ and the condition (36) ensures that $f_1(t_1)$ and $f_2(t_2)$ are negative. Furthermore, consider $t_0 = a\frac{h_2+h_3-h_2h_3}{h_2+h_3}$, which is the root of $(1 - h_2)(1 - h_3)(-h_2 - h_3)f_1(t) - f_2(t)$. It holds that $t_1 < t_0 < t_2$ and both $f_1(t_0)$ and $f_2(t_0)$ are positive, which implies that the positive intersection $f_1(t)$ and $f_2(t)$ are positive. We conclude that when $t = t_0$, the matrix $G(t_0)$ is stable, and when t is large, $G(t)$ is again stable. Yet when $t = t_2 \in (t_0, \infty)$, the matrix $G(t_2)$ is not stable.

□