Gradient dominance of a generalized framework from linear quadratic regulator problem

Yue Sun

June 20, 2020

1 Introduction

This chapter studies the optimization algorithms for a class of optimal control problems. We start from the simple linear time invariant system and state feedback control. The continuous linear system dynamics is (there is possibly a processing noise as well)

$$\dot{x} = Ax + Bu,$$

and the discrete time dynamics is

$$x(t+1) = Ax(t) + Bu(t).$$

The goal is to find the set of inputs u from time zero, that minimizes some loss form that typically depends on state and input. Probably the most famous and well studied optimal control problem is linear quadratic regulator, where we minimize a quadratic loss

$$\int_0^\infty (x(t)^T Qx(t) + u(t)^T Ru(t))dt$$

for continuous time and

$$\sum_{i=0}^{\infty} (x(t)^T Qx(t) + u(t)^T Ru(t))$$

for discrete time. When the time horizon is restricted to finite time, especially in discrete case, it's easy to write the problem as convex optimization and directly solve it [1]. For infinite time horizon, it is known that the optimal controller is linear in state, say $u(t) = Kx(t)$ for a constant K [2]. This can be extended to LQG when the output can be observed and the state can be recovered by Kalman filter [3]. One can solve the Riccati equation to obtain the optimal controller [4, 5].

It becomes more difficult when the system is not linear, where the dynamics $Ax + Bu$ is replaced by a function of state and input. The classical solution is dynamic programming, or solving Bellman equations [6]. Recently, thanks to the development of reinforcement learning theory, this method is revisited a lot, and deep learning enables it to work really well even in the highly complicated systems. But it's still mysterious how those deep learning models work, and recently people research the better known linear systems in the hope of understanding the complex ones.

Usually people do not know the dynamics beforehand when they encounter a dynamical system, and among the optimal control algorithms, there are two major types: **model based methods**, when the system is first identified and then a controller is trained, usually used when we have a good parameterization of the system; or **model free methods**, when the controller is directly trained from the loss without characterizing the dynamics. For LQR, model based methods are largely studied. System identification dates back to [7,8] etc., and recently [?,9,10] gives sample complexity bounds for state-observed system, and [11,12] for output-observed system. [?,13,14] describe the procedure of the joint system identification and optimal control based on the estimate of the system. The model free method is proposed by [15], named policy iteration, and more recently reviewed by [16,17]. It can be proven that, this approach, when applied in LQR, is a nonconvex optimization problem, and it's of interest whether we can still get a convergence guarantee for training. [18,19] proved the convergence for discrete LQR, respectively by leveraging dynamic programming or Riccati and Lyapunov equations, and obtain *gradient dominance* from the nonconvex loss landscape. [20] proves similar results for continuous LQR.

As suggested by [20], we have better understanding of the convexification methods [21–23] by change of variables. We are interested whether it helps us understanding the easiness of solving the original nonconvex optimal control problems. Unfortunately, although [20] leverages the convexification as an intermediate step, their proof is still quite specific and does not show an intuitive connection. In this chapter, we will strat from building a bridge between convexification methods and nonconvex policy gradient methods, and generalize the guarantees for more optimal control problems.

2 Review of convexification method for continuous LQR

Convexification method is widely used in controller design problems. Define a continuous time linear time invariant system

$$\dot{x} = Ax + Bu, \ x(0) = x_0, \tag{1}$$

where x is state and u is input signal, x_0 comes from an initial distribution such that $\mathbf{E}(x_0 x_0^T) \succ 0$, one considers minimizing the LQR loss

$$\min_{u(t)} f(u(t)) := \mathbf{E}_{x_0} \int_0^\infty (x(t)^T Q x(t) + u(t)^T R u(t)) dt \tag{2}$$

where Q, R are positive definite matrices. It is known that, the input signal that minimizes the loss function $f(u)$ is a state feedback controller

$$u = K^* x = -R^{-1} B P x, \tag{3}$$

$$A^T P + PA + Q - PBR^{-1} BP = 0. \tag{4}$$

Note that once we know the state feedback controller is static, we can write loss $f(u(t))$ as $loss(K)$ which is a function of K instead, and search only in static state feedback controllers.

One approach of finding K^* is to solve the Riccati equations (3,4) to get K^*. It is also of interest of solving (2) by iterative optimization algorithms, which enables us to finish early to get smaller computational complexity, implement in noisy case or when the system parameters A, B are inexactly measured, etc. In that case, one uses a reparameterization approach [20]:

$$loss(K) = h(X, Y) = \mathbf{trace}(QX + Y^T R Y X^{-1}). \tag{5}$$

Let $\mathcal{A}(X) = AX + XA, \ \mathcal{B}(Y) = BY + Y^T B^T$, they have the relation

$$\mathcal{A}(X) + \mathcal{B}(Y) + \Omega = 0 \tag{6}$$

where $\Omega = \mathbf{E}(x_0 x_0^T)$. One can construct a bijection from X, Y to K, P, and prove that, if we minimize $h(X, Y)$ subject to (6), the optimizer X^*, Y^* will map to the optimal K^*, and this problem is convex, so we can solve it by optimization algorithms.

Recently nonconvex optimization algorithms are widely used in machine learning, so it is also of interest whether we can run gradient algorithm in K space without reparameterization, which means that, we run the gradient flow

$$\dot{K} = -\eta \nabla_K loss(K), \tag{7}$$

can K converge to the optimal controller K^*? [20] answers the question by studying the mapping between the originally and reparameterized spaces, and suggests that gradient flow converges to K^* in linear rate.

This work is a generalization of [20]. We extend the approach to a far more generic sets of problems which covers the original LQR. We will prove that, for the static state feedback controller design problems in the set of problems, gradient dominance always hold. Based on that, we argue that the nonconvex optimization problem in K space can be globally solved by gradient flow.

3 Main result

Theorem 1. *We consider the problems*

$$\begin{aligned} \min_{K} \quad & loss(K), && (8a) \\ s.t., \quad & K \text{ stabilizes} && (8b) \end{aligned}$$

and

$$\begin{aligned} \min_{P,L,G} \quad & f(L, G, P), && (9a) \\ s.t., \quad & (L, G, P) \in \mathcal{S} && (9b) \end{aligned}$$

which requires the following:

1. *L, G, P are three matrices in parameterized space. P does not have to exist, i.e., $f(L, G, P) = f(L, G)$ is allowed.*

2. *L, G, P lives in a convex feasible set $\mathcal{S}$.*

3. *Cost function $f(L, G, P)$ (or $f(L, G)$ if P does not exist) is convex.*

4. *There is a bijection between K and L, G such that $\exists P$, $(L, G, P) \in \mathcal{S}$. Specifically, $K = LG^{-1}$ where $G \succeq \lambda_0 I \succ 0$.*

5. *Cost function $loss(K) = \min_{P,(L,G,P)\in\mathcal{S}} f(L, G, P) := f(L, G, \mathscr{P}(L, G))$ when K and L, G form a bijection, and $\mathscr{P}(L, G) \in \operatorname{argmin}_{P,(L,G,P)\in\mathcal{S}} f(L, G, P)$. If P does not exist, $loss(K) = f(L, G)$. Intuitively P is some epigraph of L, G.*

Then if we solve the problem by gradient flow

$$\dot{K} = -\eta_t \nabla loss(K) \tag{10}$$

then $K(t)$ converges to the global optimizer K^*, and moreover,

1. *if f is linear, the gradient satisfies*

$$\|\nabla loss(K)\| \geq C(loss(K) - loss(K^*)). \tag{11}$$

for some constant C.

2. *if f is μ-strongly convex, the gradient satisfies*

$$\|\nabla loss(K)\| \geq C(\mu(loss(K) - loss(K^*)))^{1/2}. \tag{12}$$

for some constant C.

Remark 1. *For continuous LQR, the cost function is $f(L, G, P) = \mathrm{Tr}(QG + PR)$. $\mathcal{S}$ is intersection of $\mathcal{A}(G) + \mathcal{B}(L) + \Omega = 0$, $G \succ 0$ and $[P, L^T; L, G] \succeq 0$. $K = LG^{-1}$. We can generalize it from the basic LQR, say there is a total energy budget on input, i.e. a convex constraint $\mathrm{Tr}(P) \leq p_0$ (or even $\|P\|_\psi \leq p_0$ for some norm ψ) for a positive integer p_0.*

Lemma 1. *A linear function on a convex set $\mathcal{S}$ is gradient dominant.*

Proof. Say the function is $f(x)$, the minimum is x^*, and $x - x^* = \Delta$. Let $\nabla f(x) = g$. For any non-stationary point, $f(x) = f(x^*) + g^T\Delta$. Since $\mathcal{S}$ is a convex set, $-\Delta$ belongs to the horizon of $\mathcal{S}$ at x, so there is a direction $\frac{\Delta}{\|\Delta\|}$ such that $f(x) - f(x - t\frac{\Delta}{\|\Delta\|}) > tg^T\frac{\Delta}{\|\Delta\|}$, $t \to 0$, so the norm of projected gradient $\|\mathcal{P}_{\mathcal{S}}(\nabla f(x))\| \geq g^T\frac{\Delta}{\|\Delta\|} = \frac{f(x)-f(x^*)}{\|x-x^*\|}$.

Proof of main theorem. Let K^* be the optimal K and (P^*, L^*, G^*) be the optimal point in the parameterized space. We have $loss(K^*) = f(P^*, L^*, G^*)$. Denote u as any matrix in K space, $\mathcal{P}_u$ is projection of a vector onto direction u, then

$$\nabla loss(K)^T \nabla loss(K) \geq (\mathcal{P}_u \nabla loss(K))^T \mathcal{P}_u \nabla loss(K) = \left(\frac{\nabla loss(K)[u]}{\|u\|_F}\right)^2. \tag{13}$$

At current iteration K_t, let it map to (L_t, G_t) and $P_t = \mathscr{P}(L_t, G_t)$. Note f is convex, so

$$\begin{aligned} &\nabla f(P_t, L_t, G_t)[(P_t, L_t, G_t) - (P^*, L^*, G^*)] \\ &\geq f(P_t, L_t, G_t) - f(P^*, L^*, G^*) \\ &= f(\mathscr{P}(L_t, G_t), L_t, G_t) - f(\mathscr{P}(L^*, G^*), L^*, G^*) \\ &= loss(K_t) - loss(K^*). \end{aligned} \tag{14}$$

Now we consider the directional derivative in K space. By definition,

$$\nabla loss(K)[u] = \lim_{t\to 0^+}(loss(K + tu) - loss(K))/t.$$

Let $\Delta L = L^* - L_t$, $\Delta G = G^* - G_t$, and $u = \Delta L G_t^{-1} - L_t G_t^{-1} \Delta G G_t^{-1}$. Then

$$\begin{aligned} \nabla loss(K)[u] &= \lim_{t\to 0^+}(loss(K + tu) - loss(K))/t \\ &= \lim_{t\to 0^+}(loss(L_tG_t^{-1} + t(\Delta L G_t^{-1} - L_t G_t^{-1}\Delta G G_t^{-1})) - loss(L_t G_t^{-1}))/t \\ &= \lim_{t\to 0^+}(loss((L_t + t\Delta L)(G_t + t\Delta G)^{-1}) - loss(L_t G_t^{-1}))/t \\ &= \lim_{t\to 0^+}(f(L_t + t\Delta L, G_t + t\Delta G, \mathscr{P}(L_t + t\Delta L, G_t + t\Delta G)) - f(L_t, G_t, \mathscr{P}(L_t, G_t)))/t \\ &\leq \lim_{t\to 0^+}(f(L_t + t\Delta L, G_t + t\Delta G, P_t + t\Delta P) - f(L_t, G_t, \mathscr{P}(L_t, G_t)))/t \\ &= \nabla f(P_t, L_t, G_t)[(P^*, L^*, G^*) - (P_t, L_t, G_t)] \end{aligned}$$

So

$$\nabla loss(K)[-u] \geq \nabla f(P_t, L_t, G_t)[(P_t, L_t, G_t) - (P^*, L^*, G^*)] > 0.$$

Using (13) and (14), we have

$$\nabla loss(K_t)^T \nabla loss(K_t) \geq \frac{1}{\|u\|_F^2}(loss(K_t) - loss(K^*))^2$$

Note that, although f is linear, it interacts with a general convex set $\mathcal{S}$ so that we cannot get the relation between $(\Delta L, \Delta G)$ and $f(P_t, L_t, G_t) - f(P^*, L^*, G^*)$, or equivalently with $loss(K) - loss(K^*)$. So we cannot further cancel $(loss(K_t) - loss(K^*))^2$ with $\|u\|_F^2$ despite $u \sim (\Delta L, \Delta G) \sim K - K^*$. If $f(P, L, G)$ is μ strongly convex, then we can restrict f in the line segment $(P_t, L_t, G_t) - (P^*, L^*, G^*)$ and get

$$\|\mathcal{P}_{(P_t, L_t, G_t)-(P^*, L^*, G^*)} \nabla f(P_t, L_t, G_t)\| \geq \mu^{1/2}(f(P_t, L_t, G_t) - f(P^*, L^*, G^*))^{1/2}$$

then we have that

$$\begin{aligned}
&\nabla f(P_t, L_t, G_t)[(P_t, L_t, G_t) - (P^*, L^*, G^*)] \\
&= \|\mathcal{P}_{(P_t, L_t, G_t)-(P^*, L^*, G^*)} \nabla f(P_t, L_t, G_t)\| \cdot \|(P_t, L_t, G_t) - (P^*, L^*, G^*)\| \\
&\geq \mu^{1/2}(f(P_t, L_t, G_t) - f(P^*, L^*, G^*))^{1/2} \|(P_t, L_t, G_t) - (P^*, L^*, G^*)\|.
\end{aligned}$$

then

$$\begin{aligned}
\nabla loss(K_t)^T \nabla loss(K_t) &\geq \frac{1}{\|u\|_F^2}(\nabla f(P_t, L_t, G_t)[(P_t, L_t, G_t) - (P^*, L^*, G^*)])^2 \\
&\geq \frac{\mu \|(P_t, L_t, G_t) - (P^*, L^*, G^*)\|^2}{\|u\|_F^2}(f(P_t, L_t, G_t) - f(P^*, L^*, G^*)) \\
&= \frac{\mu(\|L^* - L_t\|^2 + \|G^* - G_t\|^2 + \|P^* - P_t\|^2)}{\|(L^* - L_t)G_t^{-1} - L_t G_t^{-1}(G^* - G_t)G_t^{-1}\|_F^2}(f(P_t, L_t, G_t) - f(P^*, L^*, G^*)) \\
&\geq \frac{\mu(\|L^* - L_t\|^2 + \|G^* - G_t\|^2)}{\|(L^* - L_t)G_t^{-1} - L_t G_t^{-1}(G^* - G_t)G_t^{-1}\|_F^2}(f(P_t, L_t, G_t) - f(P^*, L^*, G^*)) \\
&\geq \frac{4\mu}{(\max\{\lambda_{min}^{-1}(G_t), \lambda_{min}^{-2}(G_t)\sigma_{\max}(L_t)\})^2}(f(P_t, L_t, G_t) - f(P^*, L^*, G^*)).
\end{aligned}$$

References

[1] J. B. Rawlings, D. Q. Mayne, and M. Diehl, *Model predictive control: theory, computation, and design.* Nob Hill Publishing Madison, WI, 2017, vol. 2.

[2] R. E. Kalman *et al.*, "Contributions to the theory of optimal control," *Bol. soc. mat. mexicana*, vol. 5, no. 2, pp. 102–119, 1960.

[3] K. J. Astrom, *Introduction to stochastic control theory.* Elsevier, 1971.

[4] R. F. Stengel, *Optimal control and estimation.* Courier Corporation, 1994.

[5] G. E. Dullerud and F. Paganini, *A course in robust control theory: a convex approach.* Springer Science & Business Media, 2013, vol. 36.

[6] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.

[7] J. A. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.

[8] L. Ljung, "System identification: theory for the user," *PTR Prentice Hall, Upper Saddle River, NJ*, pp. 1–14, 1999.

[9] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," in *Conference On Learning Theory*, 2018, pp. 439–473.

[10] T. Sarkar and A. Rakhlin, "Near optimal finite time identification of arbitrary linear dynamical systems," *arXiv preprint arXiv:1812.01251*, 2019.

[11] S. Oymak and N. Ozay, "Non-asymptotic identification of lti systems from a single trajectory," *arXiv preprint arXiv:1806.05722*, 2018.

[12] T. Sarkar, A. Rakhlin, and M. A. Dahleh, "Finite-time system identification for partially observed lti systems of unknown order," *arXiv preprint arXiv:1902.01848*, 2019.

[13] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, "On the sample complexity of the linear quadratic regulator," *arXiv preprint arXiv:1710.01688*, 2017.

[14] S. Dean, S. Tu, N. Matni, and B. Recht, "Safely learning to control the constrained linear quadratic regulator," in *2019 American Control Conference (ACC).* IEEE, 2019, pp. 5582–5588.

[15] S. J. Bradtke, B. E. Ydstie, and A. G. Barto, "Adaptive linear quadratic control using policy iteration," in *Proceedings of 1994 American Control Conference-ACC'94*, vol. 3. IEEE, 1994, pp. 3475–3479.

[16] S. M. Kakade, "A natural policy gradient," in *Advances in neural information processing systems*, 2002, pp. 1531–1538.

[17] A. Rajeswaran, K. Lowrey, E. V. Todorov, and S. M. Kakade, "Towards generalization and simplicity in continuous control," in *Advances in Neural Information Processing Systems*, 2017, pp. 6550–6561.

[18] M. Fazel, R. Ge, S. M. Kakade, and M. Mesbahi, "Global convergence of policy gradient methods for linearized control problems," *arXiv preprint arXiv:1801.05039*, 2018.

[19] J. Bu, A. Mesbahi, M. Fazel, and M. Mesbahi, "Lqr through the lens of first order methods: Discrete-time case," *arXiv preprint arXiv:1907.08921*, 2019.

[20] H. Mohammadi, A. Zare, M. Soltanolkotabi, and M. R. Jovanović, "Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem," *arXiv preprint arXiv:1912.11899*, 2019.

[21] D. Youla, H. Jabr, and J. Bongiorno, "Modern wiener-hopf design of optimal controllers–part ii: The multivariable case," *IEEE Transactions on Automatic Control*, vol. 21, no. 3, pp. 319–338, 1976.

[22] V. Kučera, "A method to teach the parameterization of all stabilizing controllers," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 6355–6360, 2011.

[23] K. Zhou, J. C. Doyle, K. Glover *et al.*, *Robust and optimal control.* Prentice hall New Jersey, 1996, vol. 40.