2020 59th IEEE Conference on Decision and Control (CDC)
Jeju Island, Republic of Korea, December 14-18, 2020

# A Data-driven Approach for Constrained Infinite-Horizon Linear Quadratic Regulation

Bo Pang and Zhong-Ping Jiang

*Abstract*— **This paper presents a data-driven algorithm to solve the problem of infinite-horizon linear quadratic regulation (LQR), for a class of discrete-time linear time-invariant systems subjected to state and control constraints. The problem is divided into a constrained finite-horizon LQR subproblem and an unconstrained infinite-horizon LQR subproblem, which can be solved directly from collected input/state data, separately. Under certain conditions, the combination of the solutions of the subproblems converges to the optimal solution of the original problem. The effectiveness of the proposed approach is validated by a numerical example.**

## I. Introduction

For the past decade, data-driven control has become a popular research topic in control theory and applications [1], [2]. In data-driven control, the target control strategy is directly synthesized using the data collected from the systems, without an intermediate explicit modeling or identification step. Thus data-driven control is deemed more suitable in cases where modeling from first principles is difficult, or implementation of system identification algorithms is both costly and time-consuming.

One important subfield of data-driven control is data-driven optimal control, where reinforcement learning (RL) and approximate/adaptive dynamic programming (ADP) techniques have played a dominant role. A variety of RL/ADP methods have been proposed to achieve optimal stabilization/tracking/disturbance rejection tasks and so on, directly from the data (see [3], [4], [5], [6] and many references therein). However, most of the optimal control problems studied there do not consider state and input constraints. Due to physical limits, quality specifications, safety concerns or the limits of the hardware, state and control constraints are common in various control engineering applications. Very often, ignorance of these constraints in the controller design phase can lead to undesirable system performance or sometimes instability.

One of the reasons leading to the above difficulty is the lack of predictive capability in most of data-driven optimal control methods. This is in sharp contrast to the model predictive control (MPC), which is a powerful tool for handling state and input constraints, thanks to its predictive capability. However, in traditional MPC an explicit model must be available, which conflicts with the features of data-driven control. Recently, the emergence of data-driven MPC [7], [8], [9], [10] has mitigated this conflict. In [11], it is shown that any input/output trajectories with finite length of a discrete-time linear time-invariant (LTI) systems lies in the linear span of a finite set of persistently exciting input/output data. This fundamental result is exploited in the data-driven MPC [7], [8], [9], [10], to realize the prediction, so that the control strategies could still be obtained directly from the data.

In this paper, the infinite-horizon linear quadratic regulation (LQR) problem of discrete-time LTI systems with state and input constraints is revisited. In [12], it is shown that the optimal state trajectory of certain finite-horizon constrained LQR problem with enough long horizon, will enter an invariant set, in which the optimal solutions of the infinite-horizon unconstrained LQR problem coincide with those of the infinite-horizon constrained LQR problem. This result is exploited and extended in our paper, to propose a simple and intuitive data-driven approach to find near-optimal solutions of the constrained infinite-horizon LQR problem. The proposed approach firstly adopts a finite-horizon constrained LQR, to bring the state into an invariant set, and then uses a near-optimal controller of the infinite-horizon unconstrained LQR problem, to regulate the state to the origin. The resulting control law converges to the optimal control law of the original constrained infinite-horizon LQR, under certain conditions. Using RL/ADP techniques and the fundamental result in [11] mentioned in the last paragraph, we demonstrate that this control law could be directly synthesized from a finite set of the input/state data.

It is worth noting that even for the most basic linear quadratic setting, the problem of data-driven infinite-horizon optimal control in the presence of state and input constraints is still an active research topic and not fully solved. A safe learning method is proposed in [13], to find the best constant linear state-feedback control gain; global stabilization is achieved in [14], with the existence of actuator saturation; RL/ADP techniques are modified in [15], to asymptotically find the optimal solutions to the unconstrained LQR without violating the constraints, by assuming the state matrix is unknown. In our method, the optimal control law among all the possible control laws satisfying the state and control constraints, is directly approached from the collected input/state data, without the exact knowledge of any system matrices.

**Notation.** $\mathbb{N}$ denotes the set of natural numbers including zero; $\mathbb{N}(a) = \{a, a+1, \cdots\}$ where $a \in \mathbb{N}$; $\mathbb{N}(a, b) = \{a, a+1, \cdots, b\}$ where $a < b < \infty$ and $a, b \in \mathbb{N}$; $\otimes$ denotes the

*This work has been supported in part by the U.S. National Science Foundation under Grants ECCS-1501044 and EPCN-1903781.

B. Pang and Z.-P. Jiang are with the Control and Networks Lab, Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, 370 Jay Street, Brooklyn, NY 11201, USA (e-mail: bo.pang@nyu.edu; zjiang@nyu.edu).

978-1-7281-7447-1/20/$31.00 ©2020 IEEE

Kronecker product. For a discrete-time signal $z : \mathbb{N} \to \mathbb{R}^n$, with slight abuse of notation, we also let $z_k = z(k)$, $k \in \mathbb{N}$. We denote by $z_{[k,k+T]}$, $T \in \mathbb{N}$, the restriction in vectorized form of signal $z$ to the interval $[k, k+T]$,

$$z_{[k,k+T]} = \begin{bmatrix} z_k^T & z_{k+1}^T & \cdots & z_{k+T}^T \end{bmatrix}^T.$$

When there is no confusion, $z_{[k,k+T]}$ is also used to denote sequence $\{z_k, \cdots, z_{k+T}\}$. $z_{[0,\infty]}$ is simply denoted as boldface letter $\mathbf{z}$. The Hankel matrix associated to $z$ is defined as

$$Z_{j,t,T} = \begin{bmatrix} z_{[j,j+t-1]} & z_{[j+1,j+t]} & \cdots & z_{[j+T-1,j+T+t-2]} \end{bmatrix},$$

where $j, t, T \in \mathbb{N}$. When $t = 1$, we simply write $Z_{j,T} = Z_{j,1,T}$.

## II. Problem Formulation and Preliminaries

For a discrete-time LTI system,

$$x_{k+1} = Ax_k + Bu_k, \tag{1}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, consider the following constrained infinite-horizon linear quadratic regulation problem:

$$(P^C) \quad \begin{aligned} J(x^0) &= \inf_{\mathbf{u},\mathbf{x}} \sum_{k=0}^{\infty} r(x_k, u_k) \\ \text{s.t. } & x_{k+1} = Ax_k + Bu_k, \quad \forall k \geq 0, \quad x_0 = x^0, \\ & \mathbf{u} \in \mathbb{U}, \quad \mathbf{x} \in \mathbb{X}, \end{aligned}$$

where $\mathbb{X} = \{\mathbf{x} | x_k \in X,\ \forall k \geq 0\}$, $\mathbb{U} = \{\mathbf{u} | u_k \in U,\ \forall k \geq 0\}$, and $r(x_k, u_k) = x_k^T Q x_k + u_k^T R u_k$. Throughout the paper, we impose the following hypotheses on (1):

(H1) $Q = Q^T > 0$, $R = R^T > 0$.
(H2) $(A, B)$ is controllable.
(H3) $X$ and $U$ are closed, bounded and convex.
(H4) $0 \in \operatorname{int} X$, $0 \in \operatorname{int} U$.
(H5) $x^0 \in X_0 \triangleq \{x^0 \in \mathbb{R}^n | \exists\ \mathbf{u} \in \mathbb{U},\ \text{s.t. } \mathbf{x} \in \mathbb{X},\ \text{and } J(x^0) < \infty\}$.

Consider also the unconstrained infinite-horizon linear quadratic regulation problem:

$$(P^U) \quad \begin{aligned} J^U(x^0) &= \min_{\mathbf{u},\mathbf{x}} \sum_{k=0}^{\infty} r(x_k, u_k) \\ \text{s.t. } & x_{k+1} = Ax_k + Bu_k, \quad \forall k \geq 0, \quad x_0 = x^0. \end{aligned}$$

The Problem $P^U$ admits the unique solution $u_k = -K^* x_k$, and $x_{k+1} = (A - BK^*)x_k$, where

$$K^* = (R + B^T P^* B)^{-1} B^T P^* A$$

and $P^* = (P^*)^T$ is the unique positive-definite solution of the algebraic Riccati equation

$$P^* = A^T P^* A - A^T P^* B (R + B^T P B)^{-1} B^T P^* A + Q.$$

In this work, we assume that system matrices $A$ and $B$ are unknown, while the constraint sets $X$ and $U$ are known. Especially, we are interested in solving the Problem $P^C$ directly from the system input/state data, without the explicit identification of the parameters in the system dynamics.

Before proceeding, we first introduce some necessary preliminaries.

**Definition 1.** *A nonempty set $Z$ is a positively invariant set for system $x_{k+1} = Ax_k$, if for any $x \in Z$, $A^k x \in Z$, $\forall k \geq 0$.*

Define set

$$X_{\max} = \left\{ x^0 | \exists\ \mathbf{u} \in \mathbb{U},\ \text{s.t. } \mathbf{x} \in \mathbb{X}, \text{and } \lim_{k\to\infty} x_k = 0 \right\}.$$

**Lemma 1** ([12, Lemma 1])**.** *$X_{\max}$ is convex, and $X_{\max} = X_0$.*

Lemma 1 implies that Hypothesis (H5) is not restrictive. $X_0$ includes all the initial states that could be stabilized without violating the constraints, given $A$, $B$, $X$, $U$.

## III. Preview of the Proposed Approach

Our method is based on the following observation. Let $K \in \mathbb{R}^{m\times n}$ be a stabilizing control gain for (1) (that is, the spectral radius of A-BK is less than one), and $P_K$ denote the unique solution of Lyapunov equation

$$A_K^T P_K A_K - P_K + Q + K^T R K = 0, \tag{2}$$

where $A_K = A - BK$. Define

$$\begin{aligned} \bar{X}_K &= \{x \in \mathbb{R}^n | x \in X, -Kx \in U\}, \\ Z_{K,c} &= \{x \in \mathbb{R}^n | x^T P_K x \leq c\}, \\ O_{K,\infty} &= \{x \in \mathbb{R}^n | (A_K)^k x \in \bar{X}_K,\ \forall k \geq 0\}, \end{aligned}$$

where $c > 0$. By (2), it is easy to know that $Z_{K,c}$ is positively invariant. Since $\bar{X}_K$ is convex, by Hypothesis (H4) there exists a sufficiently small $c$, such that $Z_{K,c} \subset \bar{X}_K$. Then $Z_{K,c}$ is said to be *admissible*. That is $Z_{K,c} \subset O_{K,\infty}$. This implies that for any $k_0 \in \mathbb{N}$, if $x_{k_0} \in Z_{K,c} \subset O_{K,\infty}$, the constraints will be automatically satisfied by subsequent motion of the close-loop system with control gain $K$ at $k \geq k_0$, as if there is no constraint at all.

The above observation indicates a simple and intuitive way to stabilize the system (1) without violating the constraints: Firstly, find a finite sequence of control inputs $\{u_k\}_{k=0}^{N-1}$, $N < \infty$, to make the state $x_N$ enter an admissible set $Z_{K,c}$; Secondly, use the stabilizing control gain $K$, to regulate the state to the origin for $k \geq N$. This suggests solving the following constrained finite-horizon LQR problem

$$(P^N) \quad \begin{aligned} J^N(x^0, F) = \min_{\substack{\{u_k\}_{k=0}^{N-1}, \\ \{x_k\}_{k=0}^{N}}} & \left\{ x_N^T F x_N + \sum_{k=0}^{N-1} r(x_k, u_k) \right\} \\ \text{s.t. } & x_{k+1} = Ax_k + Bu_k, \quad \forall k = \mathbb{N}(0, N-1), \\ & x_0 = x^0, \quad \mathbf{x} \in \mathbb{X}^N, \mathbf{u} \in \mathbb{U}^N, \end{aligned}$$

where $\mathbb{X}^N = \{\mathbf{x} | x_k \in X,\ 0 \leq k \leq N\}$, $\mathbb{U}^N = \{\mathbf{u} | u_k \in U,\ 0 \leq k \leq N-1\}$, $F = F^T \in \mathbb{R}^{n\times n}$. Let $\mathbf{u}^N$ and $\mathbf{x}^N$ denote the optimal input/state trajectory of Problem $P^N$. The method is viable due to the following theorem, which is an extension of [12, Theorem 1].

**Theorem 1.** *Given a stabilizing control gain $K$, if $F \geq 0$ in Problem $P^N$, then for any admissible $Z_{K,c} \subset O_{K,\infty}$, $\exists N < \infty$, such that $x_N^N \in Z_{K,c}$.*

*Proof.* Firstly, set $F = P^*$, and let $\mathbf{u}^{N,*}$, $\mathbf{x}^{N,*}$ denote corresponding optimal input/state trajectory of Problem $P^N$. Since $Z_{K^*,c}$ is positively invariant, if $x_N^{N,*} \notin Z_{K^*,c}$, then $x_k^{N,*} \notin Z_{K^*,c}$, $\forall k \in \mathbb{N}(0,N)$. Let $q, p$ be any real numbers such that $0 < q \leq q_m \equiv \inf_{x \notin Z_{K^*,c}} \{x^T Q x\}$, $0 < p \leq p_m \equiv \inf_{x \notin Z_{K^*,c}} \{x^T P^* x\}$. Then

$$\begin{aligned} J^N(x^0, P^*) &= (x_N^{N,*})^T P^* x_N^{N,*} + \sum_{k=0}^{N-1} r(x_k^{N,*}, u_k^{N,*}) \\ &\geq Nq + p. \end{aligned}$$

Thus $x_N^{N,*} \notin Z_{K^*,c}$ implies $J^N(x^0, P^*) \to \infty$ as $N \to \infty$, which contradicts (H5). Therefore, there exists an integer $N < \infty$, such that $x_N^{N,*} \in Z_{K^*,c}$.

Now let $F$ be any real symmetric and positive semi-definite matrix, and

$$G^N(x^0) = J^N(x^0, P^*) + \left(x_N^{N,*}\right)^T (F - P^*) x_N^{N,*}.$$

From the last paragraph, we know that $\{G^N(x^0)\}_{N=1}^{\infty}$ is uniformly bounded. This fact yields

$$0 \leq J^N(x^0, F) \leq G^N(x^0) < \infty,$$

for all $N \in \mathbb{N}(1)$. By Hypothesis (H1), there must exist $N < \infty$, such that $x_N^N \in Z_{K,c}$. □

From the proof of Theorem 1, it is not hard to obtain the following corollary.

**Corollary 1.** *If $F = P^*$ in Problem $P^N$, and $N$ is chosen such that $x_N^N \in Z_{K^*,c}$, then control law*

$$u_k = \begin{cases} u_k^N, & k \in \mathbb{N}(0, N-1) \\ -K^* x_k, & k \in \mathbb{N}(N) \end{cases} \tag{3}$$

*is the optimal control law for Problem $P^C$.*

The discussions above suggest the following data-driven approach to solve the Problem $P^C$, without the exact knowledge of system dynamics:

(S1) Find near-optimal controller $\hat{K}$ and its associated $P_{\hat{K}}$ of the Problem $P^U$, directly from the data.

(S2) Compute an admissible set $Z_{\hat{K},c} \subset \bar{X}_{\hat{K}}$.

(S3) Solve the Problem $P^N$ directly from the data with $F = P_{\hat{K}}$, for increasing value of $N$ until $x_N^N \in Z_{\hat{K},c}$.

(S4) Apply $u_k^N$ for $k \in \mathbb{N}(0, N-1)$, and $u_k = -\hat{K} x_k$ for $k \in \mathbb{N}(N)$.

**Remark 1.** *If the sets $\mathbb{X}$ and $\mathbb{U}$ are polytopes, the largest $c$ in Step (S2) can be computed analytically. See [16, Equation (11)].*

The rest of this paper is organized as follows: the details for Steps (S1) and (S3) are provided in Section IV; the verification of the proposed approach by numerical experiments can be found in Section V.

## IV. Main Results

In this section, we explain how Steps (S1) and (S3) can be achieved. Firstly, by the adaptive dynamic programming techniques [17], [18], we show in subsection IV-A that the value iteration method [19] can be used to find near-optimal solutions of the Problem $P^U$ directly from the input/state data. Secondly, inspired by the recent developments in data-driven model predictive control [7], [8], [9], [10], whose ideas originate from the pioneering work in [11], we demonstrate in subsection IV-B that the Problem $P^N$ can be transformed into an equivalent data-driven optimization problem, where no explicit knowledge of $A$ and $B$ are required. Finally, the different components are assembled to solve the Problem $P^C$ in subsection IV-C, where it is proved that if the near-optimal solutions in Step (S1) converge to their optimal values, the control law in Step (S4) will converge to the optimal control law (3).

### A. Data-driven Value Iteration for the Unconstrained Infinite-horizon LQR

The value iteration method is based on the following well-known results.

**Lemma 2** ([19, Proposition 4.4.1])**.** *For $i \in \mathbb{N}$, consider iteration*

$$\begin{aligned} P_{i+1} &= A^T P_i A - A^T P_i B K_i + Q, \\ K_i &= (B^T P_i B + R)^{-1} B^T P_i A, \end{aligned} \tag{4}$$

*with $P_0 \geq 0$. If $(A, B)$ is controllable, and $(A, D)$ is observable, where $Q = D^T D$, then*

$$\lim_{i \to \infty} P_i = P^*, \qquad \lim_{i \to \infty} K_i = K^*.$$

Lemma 2 implies that we can find the optimal solutions by iterating a difference equation, starting from an initial condition. However, system matrices $A$ and $B$ appear explicitly in (4). Next we show that (4) can be solved directly from the input/state data.

Note that

$$\begin{aligned} x_{k+1}^T P_i x_{k+1} &= (Ax_k + Bu_k)^T P_i (Ax_k + Bu_k) \\ &= x_k^T A^T P_i A x_k + 2u_k^T B^T P_i A x_k + u_k^T B^T P_i B u_k. \end{aligned}$$

By the property of Kronecker product, we have

$$\begin{aligned} &\tilde{x}_{k+1}^T \operatorname{vecs}(P_i) = \tilde{x}_k^T \operatorname{vecs}(A^T P_i A) \\ &+ (2x_k \otimes u_k)^T \operatorname{vec}(B^T P_i A) + \tilde{u}_k^T \operatorname{vecs}(B^T P_i B). \end{aligned} \tag{5}$$

where the definitions of $\tilde{x}_k$, $\operatorname{vec}(\cdot)$, and $\operatorname{vecs}(\cdot)$ can be found in [20, Notations]. Suppose input/state data $x_{[0,M]}^d$ and $u_{[0,M]}^d$, $M \in \mathbb{N}$ are available, where the superscript $d$ is used to emphasize that they are the collected data to be used for the control design. We can organize (5) for the collected data into a single linear matrix equation

$$\Phi \Theta_i = \Psi \operatorname{vecs}(P_i), \tag{6}$$

where

$$\begin{aligned} \Theta_i &= \begin{bmatrix} \operatorname{vecs}^T(\Theta_{i,1}) & \operatorname{vec}^T(\Theta_{i,2}) & \operatorname{vecs}^T(\Theta_{i,3}) \end{bmatrix}^T, \\ \Theta_{i,1} &= A^T P_i A, \quad \Theta_{i,2} = B^T P_i A, \quad \Theta_{i,3} = B^T P_i B \\ \Psi &= \begin{bmatrix} \tilde{x}_1^d & \tilde{x}_2^d & \cdots & \tilde{x}_M^d \end{bmatrix}^T, \end{aligned}$$

$$\Phi = \begin{bmatrix} \delta_0 & \delta_1 & \cdots & \delta_{M-1} \end{bmatrix}^T,$$
$$\delta_j^T = \begin{bmatrix} (\tilde{x}_j^d)^T & 2(x_j^d \otimes u_j^d)^T & (\tilde{u}_j^d)^T \end{bmatrix}.$$

Note that $\Phi$ and $\Psi$ only depend on $x^d_{[0,M]}$ and $u^d_{[0,M]}$. Thus the form of (6) suggests that $\Theta_i$ can be uniquely determined, if $\Phi$ has full column rank and $P_i$ is known. To this end, the following assumption is imposed on the data $x^d_{[0,M]}$ and $u^d_{[0,M]}$.

**Assumption 1.** *Given $M \in \mathbb{N}$, $x^d_{[0,M]}$ and $u^d_{[0,M]}$, $\Phi$ has full column rank.*

With Assumption 1, Algorithm 1 is proposed to find a near-optimal solution of Problem $P^U$.

**Theorem 2.** *Under Assumption 1, in Algorithm 1,*

$$\lim_{\bar{I}\to\infty} \hat{P}_{\bar{I}} = P^*, \qquad \lim_{\bar{I}\to\infty} \hat{K}_{\bar{I}} = K^*.$$

*Proof.* By (6) and Assumption 1, if $P_0 = \hat{P}_0 \geq 0$, then solving (4) is equivalent to solving steps 4 to 6 in Algorithm 1, i.e. $P_i = \hat{P}_i$ for all $i \in \mathbb{N}(0, \bar{I})$. Thus the convergence is obtained by Lemma 2. □

**Algorithm 1** Data-driven Value Iteration for Unconstrained LQR

1: Choose $\bar{I} \in \mathbb{N}$.
2: $i \leftarrow 0$. $\hat{P}_0 \leftarrow Q$.
3: **repeat**
4: $\quad \hat{\Theta}_i \leftarrow (\Phi^T\Phi)^{-1}\Phi^T\Psi \operatorname{vecs}(\hat{P}_i)$
5: $\quad \hat{K}_i \leftarrow (\hat{\Theta}_{i,3} + R)^{-1}\hat{\Theta}_{i,2}$
6: $\quad \hat{P}_{i+1} \leftarrow \hat{\Theta}_{i,1} + \hat{\Theta}_{i,2}^T\hat{K}_i + Q$
7: $\quad i \leftarrow i + 1$
8: **until** $i > \bar{I}$

Now suppose $\hat{K}_{\bar{I}}$ is stabilizing, we derive its associated matrix $P_{\hat{K}_{\bar{I}}}$ in (2) directly from the data. Note that

$$x^d_{k+1} = A_{\hat{K}_{\bar{I}}} x^d_k + B(\hat{K}_{\bar{I}} x^d_k + u^d_k).$$

By (2) we have

$$(x^d_{k+1})^T P_{\hat{K}_{\bar{I}}} x^d_{k+1} - (x^d_k)^T P_{\hat{K}_{\bar{I}}} x^d_k = -(x^d_k)^T(\hat{K}_{\bar{I}}^T R \hat{K}_{\bar{I}} + Q)x^d_k + (u^d_k + \hat{K}_{\bar{I}} x^d_k)^T B^T P_{\hat{K}_{\bar{I}}}(2Ax^d_k + B(u^d_k - \hat{K}_{\bar{I}} x^d_k)).$$

Similar derivations to those of (6) yield

$$\Pi \begin{bmatrix} \operatorname{vecs}(P_{\hat{K}_{\bar{I}}}) \\ \operatorname{vec}(B^T P_{\hat{K}_{\bar{I}}} A) \\ \operatorname{vecs}(B^T P_{\hat{K}_{\bar{I}}} B) \end{bmatrix} = \Gamma \operatorname{vecs}(Q + \hat{K}_{\bar{I}}^T R \hat{K}_{\bar{I}}), \quad (7)$$

where

$$\Gamma = \begin{bmatrix} \tilde{x}_0^d & \tilde{x}_1^d & \cdots & \tilde{x}_{M-1}^d \end{bmatrix}^T,$$
$$\Pi = \begin{bmatrix} \xi_0 & \xi_1 & \cdots & \xi_{M-1} \end{bmatrix}^T,$$
$$\xi_j^T = \begin{bmatrix} (\tilde{x}_j^d - \tilde{x}_{j+1}^d)^T & 2(x_j^d \otimes (u_j^d + \hat{K}_{\bar{I}} x_j^d))^T & \phi_j^T \end{bmatrix},$$
$$\phi_j^T = (u_j^d - \hat{K}_{\bar{I}} x_j^d)^T \otimes (u_j^d + \hat{K}_{\bar{I}} x_j^d)^T.$$

**Lemma 3.** *If $\hat{K}_{\bar{I}}$ is stabilizing, under Assumption 1, $\Pi$ in (7) has full column rank.*

*Proof.* The proof is analogous to those of [17, Theorem 3] and [18, Lemma 3.1.]. Thus it is omitted. □

Lemma 3 implies that $P_{\hat{K}_{\bar{I}}}$ can be obtained directly from the data by solving (7), while the existence of $\bar{I}$ such that $\hat{K}_{\bar{I}}$ is stabilizing is guaranteed by the following lemma.

**Lemma 4.** *There exists $\bar{I}_0 \in \mathbb{N}$, such that for any $\bar{I} \in \mathbb{N}(\bar{I}_0 + 1)$,*

$$\hat{P}_{\bar{I}} > 0, \quad \hat{P}_{\bar{I}+1} - \hat{P}_{\bar{I}} < Q + \hat{K}_{\bar{I}}^T R \hat{K}_{\bar{I}}, \quad (8)$$

*and $\hat{K}_{\bar{I}}$ is stabilizing.*

*Proof.* Since $P^* > 0$, by continuity, Theorem 2 and Hypothesis (H1), there exists $\bar{I}_0 \in \mathbb{N}$ such that (8) is satisfied. From the proof of Theorem 2, $P_i = \hat{P}_i$ for all $i \in \mathbb{N}(0, \bar{I})$. Thus inserting (4) into the second inequality of (8) yields

$$(A - B\hat{K}_{\bar{I}})^T P_{\bar{I}}(A - B\hat{K}_{\bar{I}}) - P_{\bar{I}} < 0.$$

By the Lyapunov lemma, $\hat{K}_{\bar{I}}$ is stabilizing. □

Condition (8) is helpful because it can be checked in Algorithm 1.

### *B. A Data-driven Method for the Constrained Finite-horizon LQR*

In this subsection, we assume horizon $N$ of the Problem $P^N$ is fixed, and sequences of input/state data $x^d_{[0,T-1]}$ and $u^d_{[0,T-1]}$ generated by system (1) are available, where $T \in \mathbb{N}(1)$.

**Definition 2** ([11])**.** *The signal $z_{[0,T-1]}$ is persistently exciting of order $L$ if its associated Hankel matrix $Z_{0,L,T-L+1}$ has full rank $\sigma L$, where $\sigma \in \mathbb{N}$ is the dimension of the signal.*

The following results are key ingredients of this subsection, where for $t \in \mathbb{N}(1)$, $X_{0,t,T-t+1}$ and $U_{0,t,T-t+1}$ are Hankel matrices associated with $x^d_{[0,T-1]}$ and $u^d_{[0,T-1]}$, respectively.

**Lemma 5** ([11, Corollary 2])**.** *If the input $u^d_{[0,T-1]}$ is persistently exciting of order $n + t$, $t \in \mathbb{N}(1)$, then*

$$\operatorname{rank} \begin{bmatrix} U_{0,t,T-t+1} \\ X_{0,T-t+1} \end{bmatrix} = n + tm. \quad (9)$$

**Lemma 6** ([11, Theorem 1])**.** *Given $t \in \mathbb{N}(1)$,*

*i) If $u^d_{[0,T-1]}$ is persistently exciting of order $n + t$, then any $t$-long input/state trajectory of system (1) can be expressed as*

$$\begin{bmatrix} u_{[0,t-1]} \\ x_{[0,t-1]} \end{bmatrix} = \begin{bmatrix} U_{0,t,T-t+1} \\ X_{0,t,T-t+1} \end{bmatrix} g,$$

*where $g \in \mathbb{R}^{T-t+1}$.*

*ii) For any $g \in \mathbb{R}^{T-t+1}$,*

$$\begin{bmatrix} U_{0,t,T-t+1} \\ X_{0,t,T-t+1} \end{bmatrix} g$$

*is a $t$-long input/state trajectory of system (1).*

Lemma 6 is originally proven in [11, Theorem 1] in the behavioral framework. It makes the replacement of the parametric description of system (1) with finite data possible. A recent proof of this result in the state-space framework can be found in [2, Lemma 2].

At first glance, it seems by Lemma 6 that only $t$-long input/state trajectory can be represented by the collected data with persistent excitation of order $t+n$. And the longer the trajectory we want to represent, the larger the order of persistent excitation of the data should be. But this is not necessary, as demonstrated in [10] for input/output systems. Actually, state/input trajectory of arbitrary finite length can be represented by collected data with a fixed order $2+n$ of persistent excitation, by weaving pieces of 2-long trajectories one after another.

**Lemma 7.** *If $u^d_{[0,T-1]}$ is persistently exciting of order $n+2$, then any state/input trajectory of system* (1) *with length $H<\infty$ can be represented by*

$$\begin{bmatrix} u_{[j,j+1]} \\ x_j \end{bmatrix} = \begin{bmatrix} U_{0,2,T-1} \\ X_{0,T-1} \end{bmatrix} g_j, \tag{10}$$

$$x_{[j,j+1]} = [X_{0,2,T-1}]\, g_j, \tag{11}$$

*for all $j \in \mathbb{N}(0,H-2)$, where $g_j \in \mathbb{R}^{T-1}$.*

*Proof.* For each $j \in \mathbb{N}(0,H-2)$, by (9) and the Rouché Capelli theorem, there exists a $g_j$ satisfying (10). Then (11) follows from

$$\begin{aligned}[X_{1,T-1}]g_j &= A\,[X_{0,T-1}]\,g_j + B\,[U_{0,T-1}]\,g_j \\ &= Ax_j + Bu_j = x_{j+1}.\end{aligned}$$

This completes the proof. □

Analogous to Lemma 6, we also have the following result.

**Lemma 8.** *For any $g_j \in \mathbb{R}^{T-1}$, $j \in \mathbb{N}(0,H-2)$, if*

$$[X_{1,T-1}]\,g_j = [X_{0,T-1}]\,g_{j+1}, \quad \forall j \in \mathbb{N}(0,H-3), \tag{12}$$

*then for each $j \in \mathbb{N}(0,H-2)$, $[U_{0,2,T-1}]\,g_j$ and $[X_{0,2,T-1}]\,g_j$ are the restrictions in vectorized form of certain state/input trajectory $u_{[0,H-1]}$ and $x_{[0,H-1]}$ of system* (1) *to the interval $[j,j+1]$.*

*Proof.* By Item ii) in Lemma 6, $[U_{0,2,T-1}]\,g_j$ and $[X_{0,2,T-1}]\,g_j$ are 2-long state/input trajectories of system (1). Condition (12) weaves these 2-long trajectories into one single $H$-long trajectory, which completes the proof. □

Using Lemmas 7 and 8, we are able to substitute system (1) with (10)–(12) in the Problem $P^N$, to obtain an optimization problem only involving collected data, without the exact knowledge of system matrices.

$(P^{N,D})$

$$J^{N,D}(x^0,F) = \min_{\{g_k\}_{k=0}^{N-1}} \left\{ x_N^T F x_N + \sum_{k=0}^{N-1} r(x_k,u_k) \right\}$$

$$\text{s.t.} \begin{bmatrix} u_{[k,k+1]} \\ x_{[k,k+1]} \end{bmatrix} = \begin{bmatrix} U_{0,2,T-1} \\ X_{0,2,T-1} \end{bmatrix} g_k,\ x_0 = x^0, \tag{13}$$

$$[X_{1,T-1}]\,g_k = [X_{0,T-1}]\,g_{k+1}, \tag{14}$$

$$\mathbf{x} \in \mathbb{X}^N, \mathbf{u} \in \mathbb{U}^N,\ \forall k \in \mathbb{N}(0,N-2).$$

In the above optimization problem, $u_{[0,N]}$ and $x_{[0,N]}$ are completely determined by the collected data $u^d_{[0,T-1]}$, $x^d_{[0,T-1]}$ and $g_k, k \in \mathbb{N}(0,N-1)$. Thus the only independent decision variables are $g_k, k \in \mathbb{N}(0,N-1)$.

**Theorem 3.** *If $u^d_{[0,T-1]}$ is persistently exciting of order $n+2$, then Problem $P^{N,D}$ is feasible, and the optimal state and control trajectories of Problem $P^{N,D}$ coincide with those of Problem $P^N$.*

*Proof.* Let $\mathcal{B}_{N,1}$ denote the set of all $N$-long input/state trajectories satisfying (1) with initial condition $x^0$. Let $\mathcal{B}_{N,2}$ denote the set of all $N$-long input/state trajectories that can be generated by (13) and (14). By Lemmas 7 and 8, $\mathcal{B}_{N,1} = \mathcal{B}_{N,2}$. Thus the feasible sets and cost functions of the two optimization problems are same. So do their optimal solutions. □

### *C. Synthesized Algorithm for the Data-driven Constrained Infinite-horizon LQR*

Finally, all the components discussed above are assembled into Algorithm 2.

**Algorithm 2** Data-driven Constrained LQR

**Input:** constraint sets $X$ and $U$, weighting matrices $Q$ and $R$, input/state data $(u^{d,1},x^{d,1})$ satisfying Assumption 1, input/state data $(u^{d,2},x^{d,2})$ with persistent excitation order $n+2$, initial state $x^0$, $\bar{I} \in \mathbb{N}$.

1: Find $\hat{K}_{\bar{I}}$ using Algorithm 1, and $P_{\hat{K}_{\bar{I}}}$ by solving (7).
2: Find an admissible $Z_{\hat{K}_{\bar{I}},c} \in \bar{X}_{\hat{K}_{\bar{I}}}$ (e.g. use Remark 1).
3: $N \leftarrow 0$.
4: **repeat**
5: $\quad N \leftarrow N+1$.
6: $\quad$ Solve Problem $P^{N,D}$ with $F = P_{\hat{K}_{\bar{I}}}$.
7: **until** $x_N^N \in Z_{\hat{K}_{\bar{I}},c}$.
8: Apply $\mathbf{u}^N$ for $k \in \mathbb{N}(0,N-1)$, and $u_k = -\hat{K}_{\bar{I}} x_k$ for $k \in \mathbb{N}(N)$.

The convergence of Algorithm 2 is presented in the following theorem.

**Theorem 4.** *As $\bar{I} \to \infty$, the control law given by Algorithm 2 converges to the optimal control law for Problem $P^C$.*

*Proof.* By Lemma 4, $\hat{K}_{\bar{I}}$ is stabilizing and $P_{\hat{K}_{\bar{I}}} > 0$ for all $\bar{I} \in \mathbb{N}(\bar{I}_0+1)$. Then Theorem 1 implies that Algorithm 2 will stop in finite steps. As $\bar{I} \to \infty$, $\hat{K}_{\bar{I}}$ and $P_{\hat{K}_{\bar{I}}}$ converge to $K^*$ and $P^*$ respectively. Then Corollary 1 and Theorem 3 complete the proof. □

**Remark 2.** *When the open-loop system* (1) *is stable, to satisfy Assumption 1 and the persistently exciting condition,*

*white noise or sum of sufficiently large number of sinusoids with different frequencies can be used as the control input during the data collection phase [21, Section 13.2]. If the constraints must not be violated during the data collection phase, sinusoidal signal is preferred. Since a stable LTI system is input-to-state stable [22, Section 4.9], the constraints can be satisfied by choosing small enough initial condition and the magnitude of the sinusoidal signal, see Section V for an example. When the open-loop system* (1) *is unstable, the situation is more complex. If suitable a priori information about the system is known, a linear state-feedback control gain (not necessarily optimal) can be derived by robust control techniques to stabilize the system, using e.g. [23].*

## V. Numerical Example

Consider system

$$x_{k+1} = \begin{bmatrix} 0.8 & 1 \\ 0 & 0.9 \end{bmatrix} x_k + \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} u_k,$$

which is open-loop stable. Let

$$X = \{x \in \mathbb{R}^2 |\ \|x\|_\infty \leq 5\}, \quad U = \{u \in \mathbb{R} |\ |u| \leq 1\}, \tag{15}$$

and $Q = I_2$, $R = 1$. In the simulation, to collect input/state data,

$$u_k = 0.1(\sin(-1.66k) + \sin(4.41k))$$

is applied to the system with initial condition $[0.1, 0.2]^T$, where the frequencies of the sinusoids are randomly sampled from interval $[-10, 10]$. Input/state data of length 7 is obtained, without violating the constraint sets (see Remark 2). It is checked that both the Assumption 1 and the requirement of a persistent excitation order 4 on the input data are satisfied. Algorithm 2 is implemented with $\bar{I} = 4000$, for different values of initial condition $x^0$. Obviously the constraints (15) are polytopes, thus an admissible $Z_{\hat{K}_{\bar{I}},c}$ is obtained by Remark 1 with $c = 2.8871$. The simulation results are summarized in Table I, where the differences of the costs given by Algorithm 2 and the true costs of Problem $P^C$ are less than $10^{-4}$. This validates Theorem 2 and Theorem 4.

TABLE I

| $x^0$ | $N$ | $J^N\left(x^0, P_{\hat{K}_{\bar{I}}}\right)$ | $J(x^0)$ | $J^U(x^0)$ |
|---|---|---|---|---|
| $[1, -1]^T$ | 2 | 2.5410 | 2.5410 | 2.5410 |
| $[-4, 3]^T$ | 3 | 32.2644 | 32.2644 | 30.4255 |
| $[-0.5, -2]^T$ | 4 | 15.6490 | 15.6490 | 11.2998 |
| $[3, 2]^T$ | 5 | 49.8544 | 49.8544 | 34.1136 |
| $[-0.5, -3]^T$ | 6 | 48.6237 | 48.6237 | 23.7368 |

## VI. Conclusion

A data-driven approach to solve the constrained infinite-horizon optimal control problem for linear discrete-time systems is proposed in this paper. Near-optimal controllers can be derived directly from a finite set of input/state data. The application of the proposed approach to a numerical example validates its feasibility. Robustness of the proposed method to external disturbance and measurement noises, is left for future work.

## References


[1] Z.-S. Hou and Z. Wang, "From model-based control to data-driven control: Survey, classification and perspective," *Information Sciences*, vol. 235, pp. 3 – 35, 2013.

[2] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2020.

[3] Y. Jiang and Z. P. Jiang, *Robust Adaptive Dynamic Programming*. Hoboken, New Jersey: Wiley, 2017.

[4] L. Buşoniu, T. de Bruin, D. Tolić, J. Kober, and I. Palunko, "Reinforcement learning for control: Performance, stability, and deep approximators," *Annual Reviews in Control*, vol. 46, pp. 8 – 28, 2018.

[5] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2042–2062, 2017.

[6] D. Wang, H. He, and D. Liu, "Adaptive critic nonlinear robust control: A survey," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3429–3451, 2017.

[7] H. Yang and S. Li, "A data-driven predictive controller design based on reduced Hankel matrix," in *Proceedings of the 10th Asian Control Conference (ASCC)*, Sabah, Malaysia, 2015, pp. 1–7.

[8] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the DeePC," in *Proceedings of the 18th European Control Conference (ECC)*, Naples, Italy, 2019, pp. 307–312.

[9] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Transactions on Automatic Control*, 2020.

[10] J. Berberich and F. Allgöwer, "A trajectory-based framework for data-driven system analysis and control," in *Proceedings of the 19th European Control Conference (ECC)*, Saint Petersburg, Russia, 2020, pp. 1365–1370.

[11] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor, "A note on persistency of excitation," *Systems & Control Letters*, vol. 54, no. 4, pp. 325–329, 2005.

[12] D. Chmielewski and V. Manousiouthakis, "On constrained infinite-time linear quadratic optimal control," *Systems & Control Letters*, vol. 29, no. 3, pp. 121–129, 1996.

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

[14] S. A. A. Rizvi and Z. Lin, "An iterative Q-learning scheme for the global stabilization of discrete-time linear systems subject to actuator saturation," *International Journal of Robust and Nonlinear Control*, vol. 29, no. 9, pp. 2660–2672, 2019.

[15] A. Chakrabarty, R. Quirynen, C. Danielson, and W. Gao, "Approximate dynamic programming for linear systems with state and input constraints," in *Proceedings of the 18th European Control Conference (ECC)*, Naples, Italy, 2019, pp. 524–529.

[16] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002.

[17] B. Pang, T. Bian, and Z. P. Jiang, "Adaptive dynamic programming for finite-horizon optimal control of linear time-varying discrete-time systems," *Control Theory and Technology*, vol. 17, no. 1, pp. 73–84, 2019.

[18] W. Gao, Y. Jiang, Z. P. Jiang, and T. Chai, "Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming," *Automatica*, vol. 72, pp. 37 – 45, 2016.

[19] D. P. Bertsekas, *Dynamic programming and optimal control*, 3rd ed. Athena scientific Belmont, MA, 2005, vol. 1.

[20] B. Pang, T. Bian, and Z.-P. Jiang, "Robust policy iteration for continuous-time linear quadratic regulation," *arXiv preprint arXiv:2005.09528*, 2020.

[21] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[22] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, New Jersey: Prentice-Hall, 2002.

[23] G. Garcia, D. Arzelier *et al.*, "Robust stabilization of discrete-time linear systems with norm-bounded time-varying uncertainty," *Systems & Control Letters*, vol. 22, no. 5, pp. 327–339, 1994.