Adaptive Control of Positive Systems with Application to Learning SSP

Fethi Bencherki FETHI.BENCHERKI@CONTROL.LTH.SE
Anders Rantzer ANDERS.RANTZER@CONTROL.LTH.SE
Department of Automatic Control, Lund University, Sweden

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

An adaptive controller is proposed and analyzed for the class of infinite-horizon optimal control problems in positive linear systems presented in (Ohlin et al., 2024b). This controller is derived from the solution of a "data-driven algebraic equation" constructed using the model-free Bellman equation from Q-learning. The equation is driven by data correlation matrices that do not scale with the number of data points, enabling efficient online implementation. Consequently, a sufficient condition guaranteeing stability and robustness to unmodeled dynamics is established. The derived results also provide a quantitative characterization of the interplay between excitation level and robustness to unmodeled dynamics. The class of optimal control problems considered here is equivalent to Stochastic Shortest Path (SSP) problems, allowing for a performance comparison between the proposed adaptive policy and model-free algorithms for learning the stochastic shortest path, as demonstrated in the numerical experiment.

Keywords: Adaptive Control, Positive Systems, Data-Driven Control

1. Introduction

Positive systems represent a class of dynamical systems in which the state remains nonnegative for all time, provided the initial condition is nonnegative. Many physical variables—such as concentrations, buffer levels, queue lengths, charge levels, and prices—are inherently nonnegative, motivating mathematical models that incorporate this structural constraint. Such models have been successfully employed in diverse application domains, including traffic and congestion modeling (Shorten et al., 2006), thermodynamics (Haddad et al., 2010), biology and pharmacology (Carson and Cobelli, 2013; Blanchini and Giordano, 2014), and epidemiology (Hernandez-Vargas and Middleton, 2013). These wide-ranging applications have spurred significant research interest in the analysis and control of positive systems. Foundational and recent contributions in this area include (De Leenheer and Aeyels, 2001; Rami and Tadeo, 2007; Rantzer, 2015; Ebihara et al., 2016; Gurpegui et al., 2023; Ohlin et al., 2024b), as well as the comprehensive tutorial (Rantzer and Valcher, 2018).

Despite the extensive literature on the optimal control of positive systems, most existing approaches remain offline and model-based. This work is motivated by recent advancements and growing interest in statistical machine learning (Tsiamis et al., 2023) and finite-time analysis, aiming to develop an online, data-driven approach to the problem. This direction has gained increasing attention in recent years, as seen in studies addressing the linear quadratic regulator (LQR) problem (De Persis and Tesi, 2019; Markovsky and Dörfler, 2021; Zhao et al., 2024) and more recent works focused on positive systems (Shafai et al., 2022; Miller et al., 2023; Padoan et al., 2023; Iwata et al., 2024; Al Makdah and Pasqualetti, 2024; Wang and Shafai, 2024).

In the current manuscript, we adopt a worst-case approach regarding the types of disturbances and uncertain plant parameters, consistent with the works presented in (Rantzer, 2021; Kjellqvist

and Rantzer, 2022a,b; Bencherki and Rantzer, 2023; Renganathan et al., 2023). However, the disturbances in this paper are subject to a bounded constraint based on past states and inputs, similar to the approach in (Rantzer, 2024), where the online linear quadratic optimal control problem in the presence of non-stochastic process disturbances is addressed. This constraint introduces a different perspective compared to the aforementioned works.

1.1. Contributions and outline of the paper

Contributions The paper proposes and analyzes an adaptive control scheme for the class of optimal control of positive systems presented in (Ohlin et al., 2024b). A data-driven algebraic equation is constructed from the model-free counterpart of the Bellman optimality equation. This equation allows for the direct extraction of the adaptive policy, bypassing an explicit system identification step. The data-driven equation is constructed and updated under the assumption of available full-state measurements contaminated with additive process noise. To account for uncertainties related to unmodeled dynamics and time-variations in the plant, we avoid making stochastic assumptions about the noise. This ensures that the proposed approach is equipped with robustness guarantees.

Applications The considered problem class can capture various network routing problem settings. A specific instance of this appeared in (Bencherki and Rantzer, 2024a), where the authors addressed the problem of learning the optimal processing rate over processing networks. Another interesting instance is the Stochastic Shortest Path (SSP) problem class, due to the work in (Ohlin et al., 2024a), where the authors demonstrate the equivalence between the two problem classes. This equivalence motivates a numerical study comparing the performance of the model-free policy presented here with existing parameter-free algorithms for SSP.

Outline The paper is structured as follows: Section 2 outlines the problem setup, starting with the model-based optimal control problem from (Ohlin et al., 2024b), progressing to its model-free counterpart, and deriving the algebraic equation for the adaptive policy. Section 3 presents supporting lemmas, leading to the formal online performance analysis of the proposed adaptive policy. Section 4 compares the numerical performance of the adaptive policy with the Q-learning algorithm from (Yu and Bertsekas, 2013) in learning the stochastic shortest path. Finally, Section 5 concludes the paper and discusses future directions.

1.2. Notation

Inequalities are applied element-wise to matrices and vectors throughout. Furthermore, the notation $\mathbb{R}^n_+$ denotes the closed positive orthant of dimension n. The symbol $|X|$ represents the element-wise absolute value of the entries of the matrix (or vector) X. The operator $\min\{A, 0\}$ extracts the minimum element of A, yielding zero if A contains no negative elements. The expressions $\mathbf{1}_{p\times q}$ and $\mathbf{0}_{p\times q}$ represent matrices of ones or zeros, respectively, of the indicated dimension, with the subscript omitted when the size is clear from the context. The operation $\otimes$ denotes the Kronecker product. $U(a, b)$ denotes the uniform distribution between a and b.

2. Problem setup

We consider the class of infinite-horizon optimal control problems presented in (Ohlin et al., 2024b):

$$\begin{aligned} \text{Minimize} \quad & \sum_{t=0}^{\infty} \left[s^\top x(t) + r^\top u(t) \right] \quad \text{over } \{u(t)\}_{t=0}^{\infty} \\ \text{subject to} \quad & x(t+1) = Ax(t) + Bu(t) \\ & u(t) \geq 0, \quad x(0) = x_0 \\ & \mathbf{1}^\top u_1(t) \;\leq\; E_1^\top x(t) \\ & \quad \vdots \qquad\qquad \vdots \\ & \mathbf{1}^\top u_n(t) \;\leq\; E_n^\top x(t) \end{aligned} \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ and $B = \begin{bmatrix} B_1 & \cdots & B_n \end{bmatrix} \in \mathbb{R}^{n\times m}$, where each $B_i \in \mathbb{R}^{n\times m_i}$, define the linear dynamics. The input signal $u \in \mathbb{R}^m$ is partitioned into n subvectors u_i, each containing m_i elements, such that $m = \sum_{i=1}^n m_i$. The cost vectors associated with the states and control inputs are $s \in \mathbb{R}^n_+$ and $r \in \mathbb{R}^m_+$, where each $r_i \in \mathbb{R}^{m_i}_+$ follows the partitioning of u. The constraints on the input signal u are given by $E = \begin{bmatrix} E_1 & \cdots & E_n \end{bmatrix}^\top \in \mathbb{R}^{n\times n}_+$. Furthermore, we define the extended constraint matrix $\bar{E} = \begin{bmatrix} \mathbf{1}^\top_{m_1} \otimes E_1 & \cdots & \mathbf{1}^\top_{m_n} \otimes E_n \end{bmatrix}^\top \in \mathbb{R}^{m\times n}_+$, and the set of indices $\mathcal{V} = \{1, \ldots, n\}$. Let $K = \begin{bmatrix} K_1^\top & \cdots & K_n^\top \end{bmatrix}^\top$ be a feedback matrix with $K_i \in \mathbb{R}^{m_i\times n}_+$, and define the set of feasible gains as

$$\mathcal{K}(E) \triangleq \left\{ K : (\forall i \in \mathcal{V})\ \mathbf{1}^\top_{m_i} K_i = E_i^\top \text{ or } \mathbf{1}^\top_{m_i} K_i = \mathbf{0}_{1\times n} \right\}. \tag{2}$$

Correspondingly, the state feedback law for the i-th control subvector is given by $u_i = K_i x$. The set $\mathcal{K}(E)$ characterizes all feedback gain matrices that result in either full or zero actuation of the control inputs u_i, for $i \in \mathcal{V}$. We impose the following two assumptions on the sextuple $(A, B, E, \bar{E}, s, r)$.

Assumption 1 ((Ohlin et al., 2024a)) *The matrices A, B and the set $\mathcal{K}(E)$ satisfy $(A+BK)x \geq 0$ for all $K \in \mathcal{K}(E)$ and all reachable states $x \in \mathbb{R}^n_+$.*

Assumption 2 *The triplet $(s, \bar{E}, r)$ satisfies $s > \bar{E}^\top r$.*

Remark 1 Assumption 1 ensures the positivity of the closed-loop dynamics. Assumption 2, on the other hand, requires that $s > 0$, which guarantees detectability of the system and ensures that the optimal feedback law is also stabilizing (Ohlin et al., 2024a).

2.1. Solution to problem 1

Under Assumption 1 and via dynamic programming (Bellman, 1966), it was shown in (Ohlin et al., 2024b, Theorem 1) that if problem (1) has a finite value for every $x(0) \in \mathbb{R}^n_+$, then the optimal cost would be $p^\top x(0)$, where $p \in \mathbb{R}^n_+$ is the solution to the following model-based algebraic equation

$$p = s + A^\top p + \sum_{i=1}^{n} \min\{r_i + B_i^\top p, 0\} E_i. \tag{3}$$

Furthermore, the optimal policy is a linear state feedback law $u(t) = Kx(t)$, where

$$K = \begin{bmatrix} K_1^\top & \ldots & K_n^\top \end{bmatrix}^\top, \quad \text{with} \quad K_i \triangleq \begin{bmatrix} \mathbf{0}_{(j-1)\times n} \\ E_i^\top \\ \mathbf{0}_{(m_i-j)\times n} \end{bmatrix} \in \mathbb{R}^{m_i\times n}, \tag{4}$$

where the vector $E_i^\top$ is placed at the j-th row, with j being the index of the minimal element of $r_i + B_i^\top p$, provided it is negative. If all elements are nonnegative, then $K_i = \mathbf{0}_{m_i \times n}$. The solution to (3) can, for instance, be obtained via value iteration (VI), by performing the following fixed-point iteration on the parameter p until convergence

$$p^{k+1} = s + A^\top p^k + \sum_{i=1}^{M} \min\{r_i + B_i^\top p^k, 0\} E_i, \quad p^0 = 0. \tag{5}$$

2.2. Model-free optimal control of positive systems via Q-factor

The cost-to-go function from time t for the optimization problem in (1) under a control policy u, starting at time t from state $x(t)$, is given by

$$J(x(t)) \triangleq \min_u \sum_{k=t}^{\infty} s^\top x(k) + r^\top u(k).$$

If finite, this optimization problem yields the objective value $J(x(t)) = p^\top x(t)$, where p is determined by solving (3). The optimal Q-function, as defined in (Bradtke et al., 1994), is

$$Q(x(t),\, u(t)) \triangleq c(x(t),\, u(t)) + J(x(t+1)), \tag{6}$$

which represents the cost of taking action $u(t)$ starting at state $x(t)$ and subsequently following the optimal policy u^*. The optimal Q-function is then given by

$$Q^*(x(t),\, u(t)) = \begin{bmatrix} s^\top + p^\top A & r^\top + p^\top B \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} = \begin{bmatrix} q^x \\ q^u \end{bmatrix}^\top \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} = q^\top \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}, \tag{7}$$

where $q \triangleq \begin{bmatrix} q^x \\ q^u \end{bmatrix} \triangleq \begin{bmatrix} s + A^\top p \\ r + B^\top p \end{bmatrix}$. It also holds that $J(x(t)) = \min_{u(t) \in \mathcal{U}(x(t))} Q(x(t),\, u(t))$, where $\mathcal{U}(x(t))$ denotes the set of inputs satisfying the constraints in (1) at time t. Then, from (6), it follows

$$Q(x(t),\, u(t)) = c(x(t),\, u(t)) + \min_{u(t+1) \in \mathcal{U}(x(t+1))} Q(x(t+1),\, u(t+1)), \tag{8}$$

and the optimal policy is given by $u^*(t) = \arg\min_{u(t) \in \mathcal{U}(x(t))} Q(x(t),\, u(t))$. One can interpret (8) as the Bellman equation in the Q-factor formulation (Sutton and Barto, 2018). By virtue of the definition in (2), replacing (7) in (8) yields

$$\begin{bmatrix} q^x \\ q^u \end{bmatrix}^\top \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} = s^\top x(t) + r^\top u(t) + \min_{K \in \mathcal{K}(E)} \begin{bmatrix} q^x \\ q^u \end{bmatrix}^\top \begin{bmatrix} I \\ K \end{bmatrix} x(t+1). \tag{9}$$

In the absence of knowledge of the dynamics (A, B), equation (9) enables us to obtain information about the q-parameter by collecting triplets $(x(t),\, u(t),\, x(t+1))$. In fact, collecting $t+1$ consecutive data points, for any $t \geq 1$, leads to

$$\left(q - \begin{bmatrix} s \\ r \end{bmatrix} \right)^\top \begin{bmatrix} x(0) & \cdots & x(t-1) \\ u(0) & \cdots & u(t-1) \end{bmatrix} = \min_{K \in \mathcal{K}(E)} q^\top \begin{bmatrix} I \\ K \end{bmatrix} \begin{bmatrix} x(1) & \cdots & x(t) \end{bmatrix}. \tag{10}$$

Multiplying equation (10) from the right by $\begin{bmatrix} \lambda^{t-1}x(0) & \lambda^{t-2}x(1) & \cdots & x(t-1) \\ \lambda^{t-1}u(0) & \lambda^{t-2}u(1) & \cdots & u(t-1) \end{bmatrix} \in \mathbb{R}^{(n+m)\times t}$ for a forgetting factor $\lambda \in (0,\, 1]$, and defining the data correlation matrices

$$\Sigma(t) \triangleq \sum_{k=0}^{t-1} \lambda^{t-1-k} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^\top + \lambda^t \Sigma(0) \;\text{ and }\; \bar{\Sigma}(t) \triangleq \sum_{k=0}^{t-1} \lambda^{t-1-k} x(k+1) \begin{bmatrix} x(k) \\ u(k) \end{bmatrix}^\top, \tag{11}$$

yields the *data-driven algebraic equation* in the $q(t)$-parameter

$$\left(q(t) - \begin{bmatrix} s \\ r \end{bmatrix} \right)^\top \Sigma(t) = \min_{K(t)\in\mathcal{K}(E)} q^\top(t) \begin{bmatrix} I \\ K(t) \end{bmatrix} \bar{\Sigma}(t). \tag{12}$$

Here, we denote the data-based solution by $(q(t), K(t))$, in contrast to the model-based (or ground-truth) solution (q, K). Equation (12) forms the foundation for the construction of the proposed controller, as will be shown in the sequel.

2.3. Problem Formulation

Inspired by (12), for the linear system

$$x(t+1) = Ax(t) + Bu(t) + w(t), \tag{13}$$

we propose and analyze the performance of policies of the form

$$\begin{cases} \Sigma(t) = \lambda\Sigma(t-1) + \left[x^\top(t-1) \;\; u^\top(t-1) \right]^\top \left[x^\top(t-1) \;\; u^\top(t-1) \right], \quad \Sigma(0) \succ 0, \\ \bar{\Sigma}(t) = \lambda\bar{\Sigma}(t-1) + x(t) \left[x^\top(t-1) \;\; u^\top(t-1) \right], \quad \bar{\Sigma}(0) = 0, \\ u(t) = K(t)x(t) + \epsilon(t). \end{cases} \tag{14}$$

The controller states $\Sigma(t)$ and $\bar{\Sigma}(t)$ accumulate correlation data using a forgetting factor λ. Based on these statistics, the control gain $K(t)$ is computed as the minimizing argument of the data-driven algebraic equation in (12). To ensure sufficient excitation for learning the true system dynamics (A, B), additive exploration noise $\epsilon(t)$ is introduced.

Definition 1 *Let the model parameter set $\mathcal{M}_\beta$ be the set of plants (A, B, E) satisfying Assumptions 1 and 2, such that the algebraic equation in (3) admits a solution p satisfying*

$$s \le p \le (\beta \min_i s_i)\mathbf{1} \le \beta s.$$

Remark 2 The parameter β reflects the degree of stabilizability of the system. Larger values of β correspond to systems that are harder to stabilize.

Constrained to the model set defined in Definition 1, we aim to establish guarantees for the stability and robustness of the closed-loop system under the policy given in (14), assuming certain properties of the disturbance w. Before proceeding with the analysis, we first present methods for solving the algebraic data-driven equation in (12), as its solution is central to both the construction and implementation of the proposed controller.

2.4. Solution to data-driven algebraic equation in the $q(t)$-parameter

Similarly to the p-parameter-based algebraic equation in (3), the equation in (12) can be solved using value iteration, policy iteration, or linear programming. For brevity, policy iteration is omitted.

Solution via value iteration The solution is obtained by performing the following fixed-point iteration on the $q(t)$-parameter, starting from $q^0(t) = 0$, and continuing until convergence:

$$q^{k+1}(t) = \Sigma^{-1}(t)\bar{\Sigma}^\top(t)\begin{bmatrix} I & \left(K^k(t)\right)^\top \end{bmatrix} q^k(t) + \begin{bmatrix} s \\ r \end{bmatrix}, \quad q^0(t) = 0, \tag{15}$$

$$K_i^k(t) = \begin{cases} \begin{bmatrix} \mathbf{0}_{j-1\times n} \\ E_i^\top \\ \mathbf{0}_{m_k-j\times n} \end{bmatrix}, & \text{if } \min\left\{(q_i^u)^k(t),\, 0\right\} < 0 \\ \mathbf{0}_{m_i\times n}, & \text{otherwise} \end{cases}, \quad i = 1,\dots,n,$$

where j denotes the index of the most negative entry in $(q_i^u)^k(t)$ and the full controller matrix is $K^k(t) = \begin{bmatrix} \left(K_1^k(t)\right)^\top & \dots & \left(K_n^k(t)\right)^\top \end{bmatrix}^\top$.

Solution via linear programming Instead of using a fixed-point iteration, the solution can also be obtained via a linear programming (LP) formulation. To this end, define $z \triangleq \begin{bmatrix} (z^x)^\top & (z^u)^\top \end{bmatrix}^\top$, where $z^x \in \mathbb{R}_+^n$ and $z^u \in \mathbb{R}_+^m$. We propose the following optimization problem to solve for $q(t)$:

$$\begin{aligned} \text{maximize} \quad & \mathbf{1}^\top \begin{bmatrix} I & -E^\top \end{bmatrix} z \quad \text{over } z \in \mathbb{R}_+^{n+m},\ q(t) \in \mathbb{R}^{n+m} \\ \text{subject to} \quad & \Sigma^{-1}(t)\,\bar{\Sigma}^\top(t)\begin{bmatrix} I & -E^\top \end{bmatrix} z = q(t) - \begin{bmatrix} s \\ r \end{bmatrix}, \\ & z^x = q^x(t), \quad z^u \geq q^u(t), \quad z^u \geq 0. \end{aligned}$$

Remark 3 When the plant (A, B) is known, the term $\Sigma^{-1}(t)\,\bar{\Sigma}^\top(t)$ is replaced by the model $\begin{bmatrix} A & B \end{bmatrix}^\top$. A corresponding model-based version of the fixed-point iteration in (15) becomes:

$$q^{k+1} = \begin{bmatrix} A & B \end{bmatrix}^\top \begin{bmatrix} I & \left(K^k\right)^\top \end{bmatrix} q^k + \begin{bmatrix} s \\ r \end{bmatrix}, \quad q^0 = 0, \tag{16}$$

$$K_i^k = \begin{cases} \begin{bmatrix} \mathbf{0}_{j-1\times n} \\ E_i^\top \\ \mathbf{0}_{m_k-j\times n} \end{bmatrix}, & \text{if } \min\left\{(q_i^u)^k,\, 0\right\} < 0 \\ \mathbf{0}_{m_i\times n}, & \text{otherwise} \end{cases}, \quad i = 1,\dots,n,$$

where $K^k = \begin{bmatrix} \left(K_1^k\right)^\top & \dots & \left(K_n^k\right)^\top \end{bmatrix}^\top$, and j is the index of the most negative element in $(q_i^u)^k$. Here, the subscript t is omitted from q because the time-dependent and optimal q-parameters coincide.

3. Main results

We begin by establishing an important supporting result.

Lemma 1 *Iterating on q in (16) is algebraically equivalent to iterating on p in (5).*

Proof. The proof appears in (Bencherki and Rantzer, 2024b, Appendix A.1). ■

Remark 4 We extract two key observations from Lemma 1:

(i) Lemma 1 asserts that value iteration in the q-parameter is algebraically equivalent to value iteration in the p-parameter.

(ii) According to Lemma 1, given $(q(t), K(t))$, the solution to the data-driven algebraic equation in (12), if we define $p(t) \triangleq \begin{bmatrix} I & (K(t))^\top \end{bmatrix} q(t)$, then $p(t)$ satisfies the following model-based algebraic equation under the estimated dynamics $(\hat{A}(t), \hat{B}(t))$, provided that persistently exciting data is collected. Specifically, defining $[\hat{A}(t) \;\; \hat{B}(t)] \triangleq \bar{\Sigma}(t)\Sigma^{-1}(t)$, we obtain

$$p(t) - s = \hat{A}^\top(t)p(t) + \sum_{i=1}^{n} \min\{r_i + \hat{B}_i^\top(t)p(t), 0\}E_i. \tag{17}$$

This is key in establishing proofs for the main results of the paper, as shall be seen in Section 3.1.

Lemma 2 *Let $p \in \mathbb{R}^n_+$ and $\hat{p} \in \mathbb{R}^n_+$ be such that they satisfy the following algebraic equation and inequality, respectively, i.e., they satisfy*

$$p = s + A^\top p + \sum_{i=1}^{n} \min\{r_i + B_i^\top p, 0\}E_i \quad \textit{and} \quad \hat{p} \geq s + A^\top \hat{p} + \sum_{i=1}^{n} \min\{r_i + B_i^\top \hat{p}, 0\}E_i.$$

Then, it holds that $\hat{p} \geq p$.

Proof. The proof appears in (Bencherki and Rantzer, 2024b, Appendix A.2). ■

3.1. Performance analysis

In anticipation of our first main result, we define

$$\tilde{\Sigma}(t) \triangleq \begin{bmatrix} \Sigma^{wx}(t) & \Sigma^{wu}(t) \end{bmatrix} \triangleq \sum_{k=0}^{t-1} \lambda^{t-1-k} w(k) \begin{bmatrix} x^\top(k) & u^\top(k) \end{bmatrix},$$

which implies that $\bar{\Sigma}(t) = \begin{bmatrix} A & B \end{bmatrix} \Sigma(t) + \tilde{\Sigma}(t)$. Therefore,

$$\begin{bmatrix} \hat{A}(t) & \hat{B}(t) \end{bmatrix} = \bar{\Sigma}(t)\Sigma^{-1}(t) = \begin{bmatrix} A & B \end{bmatrix} + \begin{bmatrix} \tilde{A}(t) & \tilde{B}(t) \end{bmatrix}, \tag{18}$$

where $\begin{bmatrix} \tilde{A}(t) & \tilde{B}(t) \end{bmatrix} \triangleq \tilde{\Sigma}(t)\Sigma^{-1}(t) = \begin{bmatrix} \Sigma^{wx}(t)\Sigma^{-1}(t) & \Sigma^{wu}(t)\Sigma^{-1}(t) \end{bmatrix}$ represents the model misspecification. The first main result of the paper is stated next.

Theorem 1 *Consider β, $\rho \in \mathbb{R}_+$ satisfying $\rho\beta < 1$, and let $\Sigma(t)$ and $\bar{\Sigma}(t)$ be as in (11). Let $(A, B) \in \mathcal{M}_\beta$. Additionally, suppose that*

$$\left\|E^\top\right\|_\infty \left\|A^\top - \begin{bmatrix} I & 0 \end{bmatrix} \Sigma^{-1}(t)\bar{\Sigma}^\top(t)\right\|_\infty + \left\|B^\top - \begin{bmatrix} 0 & I \end{bmatrix} \Sigma^{-1}(t)\bar{\Sigma}^\top(t)\right\|_\infty \leq \rho, \;\; \forall t \geq t_0. \tag{19}$$

Let $p \in \mathbb{R}^n_+$ be the solution to the model-based algebraic equation in (3), and let $q(t)$ be the solution of the data-based algebraic equation in (12) with $K(t)$ being the minimizing argument. Define $p(t) \triangleq \begin{bmatrix} I & K^\top(t) \end{bmatrix} q(t)$. Then, it holds that

$$\hat{\alpha}p \leq p(t) \leq \check{\alpha}^{-1}p \tag{20}$$

for positive constants satisfying $\check{\alpha} = 1 - \rho\beta$ and $\hat{\alpha} = 1 - \check{\alpha}^{-1}\rho\beta$.

Proof. The proof appears in (Bencherki and Rantzer, 2024b, Appendix A.3). ■

Remark 5 The result in theorem 1 is obtained from a perturbation analysis to the solution of the algebraic equations in (3) and (17). Perturbation analysis of algebraic Riccati equations is a well studied problem in the literature, see the works (Konstantinov et al., 1993; Sun, 1998; Konstantinov et al., 2003).

Assumption (19) can be expressed as $\|E^\top\|_\infty \left\|\Sigma^{-1}(t)\left(\Sigma^{wx}(t)\right)^\top\right\|_\infty + \left\|\Sigma^{-1}(t)\left(\Sigma^{wu}(t)\right)^\top\right\|_\infty \leq \rho$ and holds as long as the following two conditions are satisfied:

(i) The condition number of $\Sigma(t)$ must not be too large, which depends on the choice of a suitable exploration signal $\epsilon(t)$ and a sufficiently large t_0 to ensure sufficient data is collected.

(ii) The disturbance sequence $w(t)$ remains sufficiently small relative to $x(t)$. If $w(t)$ represents unmodeled dynamics, this implies that such dynamics are subject to a gain bound from the state to the disturbance.

In cases where the disturbance sequence w is modeled as i.i.d. Gaussian noise, increasing t_0 typically enables Assumption (19) to hold for smaller values of ρ. However, if the system is exposed to adversarial disturbances, the true dynamics may never be accurately learned, as the adversary w could adopt a policy that persistently misleads the controller.

Theorem 2 *Consider $\beta,\ \rho \geq 0$ satisfying $\rho\beta < 1$, and let $\Sigma(t)$ and $\bar{\Sigma}(t)$ be as defined in (11). Let $(A, B) \in \mathcal{M}_\beta$, and let $p \in \mathbb{R}^n_+$ denote the solution to the model-based algebraic equation in (3). Suppose that (19) holds for all $t \geq t_0$, and let $K(t)$ denote the minimizer in (12). Then, it holds that*

$$\breve{\alpha}^{-1}\left(1 + \rho\beta\left(1 + \beta\left\|A + |B|\,\bar{E}\right\|_1\right)\right)p - s \geq A^\top p + K^\top(t)\left(r + B^\top p\right), \tag{21}$$

where the constant $\breve{\alpha}$ is as defined in Theorem 1.

Proof. The proof appears in (Bencherki and Rantzer, 2024b, Appendix A.4). ■

Remark 6 Theorem 2 aims to characterize how the storage function $p^\top x$ is affected when the data-driven gain $K(t)$ is used in place of the optimal gain K. As $\rho \to 0$, we have $\breve{\alpha} \to 1$, and the right-hand side of (21) approaches 1, thereby closing the suboptimality gap. Larger values of β correspond to systems that are more difficult to stabilize, necessitating smaller values of ρ to satisfy the condition $\rho\beta < 1$, which in turn reflects the need for higher-quality data.

Corollary 1 *Consider $(A,\ B) \in \mathcal{M}_\beta$, with p being the positive solution to the algebraic equation (3), and let $\beta,\ \rho > 0$ satisfy $\rho\beta < 1$. Let the system evolve as*

$$x(t+1) = Ax(t) + Bu(t) + w(t), \qquad u(t) = K(t)x(t) + \epsilon(t),$$

where $\Sigma(t)$ and $\bar{\Sigma}(t)$ are defined as in (11), and assume that (19) holds for all $t \geq t_0$. Let $q(t)$ be the solution to the data-based algebraic equation in (12), with $K(t)$ being its minimizing argument. Then, the following inequality holds:

$$\sum_{t=t_0}^{T-1} s^\top x(t) + r^\top K(t)x(t) \leq \gamma^{-1}\left(p^\top x_{t_0} + \sum_{t=t_0}^{T-1} \beta\, s^\top\,|B\epsilon(t) + w(t)|\right),$$

where $0 < \gamma \leq 1$ *satisfies* $\gamma\left(s - \bar{E}^\top |r|\right) \leq s - \bar{E}^\top |r| - \left(\breve{\alpha}^{-1} - 1 + \rho\beta\left(1 + \beta\left\|A + |B|\,\bar{E}\right\|_1\right)\right)\beta s$, *and the constant* $\breve{\alpha}$ *is as defined in Theorem 1.*

Proof. The proof appears in (Bencherki and Rantzer, 2024b, Appendix A.5). ■

Remark 7 Corollary 1 quantifies the suboptimality in the incurred cost when applying the adaptive controller in place of the optimal one. The constant γ reflects the impact of model misspecification and satisfies $\gamma \to 1$ as $\rho \to 0$, thereby tightening the suboptimality bound. The final term on the right-hand side accounts for the accumulated cost due to external disturbances $w(t)$ and the exploration noise $\epsilon(t)$.

4. Numerical experiment

We consider the problem of learning the following stochastic shortest path

Stochastic Shortest Path (SSP) problem

$$\mathcal{T}(1) = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \\ 0.4 & 0.4 \\ 0.2 & 0.2 \end{bmatrix}, \mathcal{T}(2) = \begin{bmatrix} 0 & 0.3 & 0 \\ 0.6 & 0 & 0.1 \\ 0.4 & 0.7 & 0.4 \\ 0 & 0 & 0.5 \end{bmatrix}, \mathcal{T}(3) = \begin{bmatrix} 0 & 0.2 \\ 0 & 0.2 \\ 0.4 & 0 \\ 0.6 & 0.6 \end{bmatrix}, \mathcal{T}(4) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$c(1) = \begin{bmatrix} 1.5 & 2 \end{bmatrix}, c(2) = \begin{bmatrix} 1.5 & 2 & 2 \end{bmatrix}, c(3) = \begin{bmatrix} 1.5 & 2 \end{bmatrix} c(4) = 0,\ i_{\text{init}} = 1$$

Conversion according to (Ohlin et al., 2024a)

Reformulation as problem 1

$n = 3$, $m = 4$, $x_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$ and dynamics

$$A = \begin{bmatrix} 0.4 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0.4 & 0.4 & 0.4 \end{bmatrix}, B = \begin{bmatrix} -0.4 & 0.3 & 0 & 0.2 \\ 0.4 & -0.6 & -0.5 & 0.2 \\ 0 & 0.3 & 0 & -0.4 \end{bmatrix}$$

with associated costs

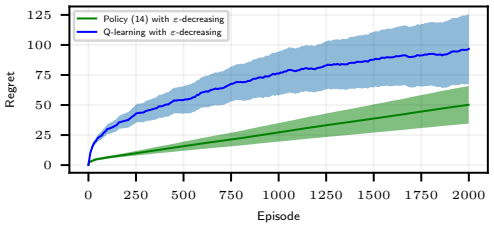
$$s = \begin{bmatrix} 1.5 & 1.5 & 1.5 \end{bmatrix}^\top, \quad r = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}^\top$$
$$E = I,\ m_1 = 1,\ m_2 = 2,\ m_3 = 1$$

where $\mathcal{T}(i)$, for $i \in \mathcal{V} \cup \{v_g\}$ with $v_g = 4$ denoting the fictitious goal state, represent the transition probabilities from state i to other states, where different columns correspond to different actions. The vector $c(i)$ denotes the expected stage cost associated with transitioning from state i to other states under different actions. Note that both Assumptions 1 and 2 hold true for our system. We compare the performance of the adaptive policy to that of the Q-learning algorithm presented in (Yu and Bertsekas, 2013), employing an ϵ-decreasing exploration strategy with $\epsilon = 0.05\alpha^h$, where $\alpha = 0.99$ and h denotes the episode number. In the context of Problem 1, ϵ-greedy exploration means selecting a random gain $K \in \mathcal{K}(E)$ with probability ϵ, and choosing the estimated optimal gain

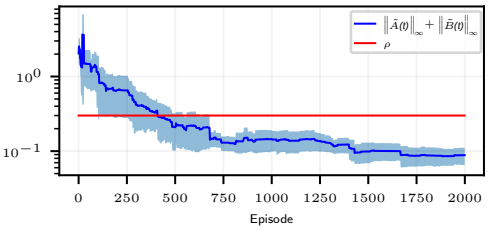
$K(t)$, computed via the data-driven algebraic equation (12), with probability $1-\epsilon$. For the policy in (14), we set $\lambda = 1$ and initialize $\Sigma(0) = 10^{-6}I$. To make the learning task more challenging, we corrupt the state measurements with additive positive disturbances $w(t) \sim U(0,\ 0.01)$ for all t. The comparison between the two algorithms is performed by evaluating the regret, defined as

$$R(H) = \left(\sum_{h=0}^{H-1} \sum_{t=0}^{T_h-1} s^\top x(t) + r^\top u(t) \right) - HJ(x(0)),$$

where H denotes the number of episodes, and T_h is the duration of episode h. The term $J(x(0))$ represents the optimal cumulative cost obtained by applying the optimal policy to the system subject to disturbances w. In the SSP domain, an episode terminates when the goal state ($i = 4$) is reached, whereas in Problem 1, an episode ends when the state vector becomes sufficiently small, indicating that the measurements are dominated by noise. After each episode, the system states are re-initialized in both problem domains. The algorithms are then restarted, retaining the latest estimates: the Q-factor for the Q-learning algorithm, and the $q(t)$-parameter and $\Sigma(t)$ for the policy in (14). For more details on the implementation, refer to the code[1]. The resulting regrets are shown in Figure 1(*a*). Both algorithms exhibit sublinear regret, with our proposed algorithm consistently outperforming the Q-learning approach. Figure 1(*b*) further illustrates a test of the condition in (19), using $\rho = 0.3$. Notably, the condition is satisfied after a sufficient number of episodes, suggesting that more accurate estimates of the system are obtained as additional data is gathered.



(*a*) Accumulated regret of each algorithm with ϵ-decreasing explorations where $\epsilon = 0.05\alpha^h$ for $\alpha = 0.99$ and h the episode number.



(*b*) The condition in (19) evaluated for $\rho = 0.3$ in the presence of disturbances. For this example, $\|E^\top\|_\infty = 1$.

Figure 1: Each plot represents the average over 100 repeated runs, with the shaded area indicating the 95% confidence interval.

5. Conclusion

The paper presented a robust adaptive data-driven control framework tailored towards the class of positive systems presented in (Ohlin et al., 2024b). This was achieved via the construction of a data-driven algebraic equation in the Q-factor, based on which the controller policy is updated in an online fashion. The designed policy proved to robustly stabilize the set $\mathcal{M}_\beta$ with robustness meant to be tolerance to a certain degree of unmodeled dynamics. The considered class witnesses applications in network routing problems among which are Stochastic Shortest Path problems, allowing for performance comparison with the existing model-free methods of finding the stochastic shortest path. Future work concerns exploring better exploration strategies than ϵ-greedy, and the possibility of adapting efficient exploration methods from the SSP literature into our control setup.

1. Code available at: https://github.com/Fethi-Bencherki/adaptive-control-positive-systems-l4dc2025

Acknowledgments

The authors thank David Ohlin at Lund University for his insightful feedback, which helped improve the paper. They are members of the ELLIIT Strategic Research Area at Lund University. This project received funding from the European Research Council (ERC) under grant agreement No. 834142 (ScalableControl) and was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation.

References

Abed AlRahman Al Makdah and Fabio Pasqualetti. Model-based and data-based output feedback for external positivity. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, pages 4119–4124, 2024. doi: 10.1109/CDC56724.2024.10885918.

Richard Bellman. Dynamic programming. *science*, 153(3731):34–37, 1966.

Fethi Bencherki and Anders Rantzer. Robust simultaneous stabilization via minimax adaptive control. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 2503–2508. IEEE, 2023. doi: 10.1109/CDC49753.2023.10384134.

Fethi Bencherki and Anders Rantzer. Data-driven adaptive dispatching policies for processing networks. *IEEE Control Systems Letters*, 8:2841–2846, 2024a. doi: 10.1109/LCSYS.2024.3516637.

Fethi Bencherki and Anders Rantzer. Adaptive control of positive systems with application to learning SSP. *arXiv preprint arXiv:2412.17012*, 2024b. URL https://arxiv.org/abs/2412.17012.

Franco Blanchini and Giulia Giordano. Piecewise-linear lyapunov functions for structural stability of biochemical networks. *Automatica*, 50(10):2482–2493, 2014. URL https://www.sciencedirect.com/science/article/pii/S0005109814003288.

Steven J Bradtke, B Erik Ydstie, and Andrew G Barto. Adaptive linear quadratic control using policy iteration. In *Proceedings of 1994 American Control Conference-ACC'94*, volume 3, pages 3475–3479. IEEE, 1994. doi: 10.1109/ACC.1994.735224.

Ewart Carson and Claudio Cobelli. *Modelling methodology for physiology and medicine*. Newnes, 2013. URL https://doi.org/10.1016/C2012-0-06031-0.

Patrick De Leenheer and Dirk Aeyels. Stabilization of positive linear systems. *Systems & control letters*, 44(4):259–271, 2001. URL https://www.sciencedirect.com/science/article/pii/S0167691101001463.

Claudio De Persis and Pietro Tesi. Formulas for data-driven control: Stabilization, optimality, and robustness. *IEEE Transactions on Automatic Control*, 65(3):909–924, 2019. doi: 10.1109/TAC.2019.2959924.

Yoshio Ebihara, Dimitri Peaucelle, and Denis Arzelier. Analysis and synthesis of interconnected positive systems. *IEEE Transactions on Automatic Control*, 62(2):652–667, 2016. doi: 10.1109/TAC.2016.2558287.

Alba Gurpegui, Emma Tegling, and Anders Rantzer. Minimax linear optimal control of positive systems. *IEEE Control Systems Letters*, 2023. doi: 10.1109/LCSYS.2023.3341344.

Wassim M Haddad, VijaySekhar Chellaboina, and Qing Hui. *Nonnegative and compartmental dynamical systems*. Princeton University Press, 2010. URL https://www.jstor.org/stable/j.ctt7t21q.

Esteban A Hernandez-Vargas and Richard H Middleton. Modeling the three stages in hiv infection. *Journal of theoretical biology*, 320:33–40, 2013. URL https://www.sciencedirect.com/science/article/pii/S0022519312006170.

Takumi Iwata, Shun-ichi Azuma, Ryo Ariizumi, and Toru Asai. Data informativity for distributed positive stabilization. *IEEE Control Systems Letters*, 2024. doi: 10.1109/LCSYS.2024.3404219.

Olle Kjellqvist and Anders Rantzer. Learning-enabled robust control with noisy measurements. In *Learning for Dynamics and Control Conference*, pages 86–96. PMLR, 2022a. URL https://proceedings.mlr.press/v168/kjellqvist22a.html.

Olle Kjellqvist and Anders Rantzer. Minimax adaptive estimation for finite sets of linear systems. In *2022 American Control Conference (ACC)*, pages 260–265. IEEE, 2022b. doi: 10.23919/ACC53348.2022.9867474.

Michail M Konstantinov, P Hr Petkov, and Nikolai D Christov. Perturbation analysis of the discrete riccati equation. *Kybernetika*, 29(1):18–29, 1993.

Mihail Konstantinov, D Wei Gu, Volker Mehrmann, and Petko Petkov. *Perturbation theory for matrix equations*. Gulf Professional Publishing, 2003.

Ivan Markovsky and Florian Dörfler. Behavioral systems theory in data-driven analysis, signal processing, and control. *Annual Reviews in Control*, 52:42–64, 2021. URL https://www.sciencedirect.com/science/article/pii/S1367578821000754.

Jared Miller, Tianyu Dai, Mario Sznaier, and Bahram Shafai. Data-driven control of positive linear systems using linear programming. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 1588–1594. IEEE, 2023. doi: 10.1109/CDC49753.2023.10383859.

David Ohlin, Anders Rantzer, and Emma Tegling. Heuristic search for linear positive systems. *arXiv preprint arXiv:2410.17220*, 2024a.

David Ohlin, Emma Tegling, and Anders Rantzer. Optimal control of linear cost networks. *European Journal of Control*, page 101068, 2024b. URL https://www.sciencedirect.com/science/article/pii/S0947358024001286.

Alberto Padoan, Florian Dörfler, and John Lygeros. Data-driven representations of conical, convex, and affine behaviors. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 596–601. IEEE, 2023. doi: 10.1109/CDC49753.2023.10383687.

M Ait Rami and Fernando Tadeo. Controller synthesis for positive linear systems with bounded controls. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 54(2):151–155, 2007. doi: 10.1109/TCSII.2006.886888.

Anders Rantzer. On the kalman-yakubovich-popov lemma for positive systems. *IEEE Transactions on Automatic Control*, 61(5):1346–1349, 2015. doi: 10.1109/TAC.2015.2465571.

Anders Rantzer. Minimax adaptive control for a finite set of linear systems. In *Learning for Dynamics and Control*, pages 893–904. PMLR, 2021. URL https://proceedings.mlr.press/v144/rantzer21a.html.

Anders Rantzer. A data-driven Riccati equation. In *Proceedings of the 6th Annual Learning for Dynamics & Control Conference*, volume 242 of *Proceedings of Machine Learning Research*, pages 504–513. PMLR, 15–17 Jul 2024. URL https://proceedings.mlr.press/v242/rantzer24a.html.

Anders Rantzer and Maria Elena Valcher. A tutorial on positive systems and large scale control. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 3686–3697. IEEE, 2018. doi: 10.1109/CDC.2018.8618689.

Venkatraman Renganathan, Andrea Iannelli, and Anders Rantzer. An online learning analysis of minimax adaptive control. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 1034–1039. IEEE, 2023.

Bahram Shafai, Anahita Moradmand, and Milad Siami. Data-driven positive stabilization of linear systems. In *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)*, volume 1, pages 1031–1036. IEEE, 2022. doi: 10.1109/CoDIT55151.2022.9804005.

Robert Shorten, Fabian Wirth, and Douglas Leith. A positive systems model of tcp-like congestion control: asymptotic results. *IEEE/ACM transactions on networking*, 14(3):616–629, 2006. doi: 10.1109/TNET.2006.876178. URL https://doi.org/10.1109/TNET.2006.876178.

Ji-Guang Sun. Perturbation theory for algebraic riccati equations. *SIAM Journal on Matrix Analysis and Applications*, 19(1):39–65, 1998. doi: 10.1137/S0895479895291303. URL https://doi.org/10.1137/S0895479895291303.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Anastasios Tsiamis, Ingvar Ziemann, Nikolai Matni, and George J Pappas. Statistical learning theory for control: A finite-sample perspective. *IEEE Control Systems Magazine*, 43(6):67–97, 2023. doi: 10.1109/MCS.2023.3310345.

Yueyang Wang and Bahram Shafai. Data-driven identification and control of positive systems. In *Integrated Systems: Data Driven Engineering*, pages 289–307. Springer, 2024.

Huizhen Yu and Dimitri P Bertsekas. On boundedness of q-learning iterates for stochastic shortest path problems. *Mathematics of Operations Research*, 38(2):209–227, 2013. URL http://www.jstor.org/stable/24540850.

Feiran Zhao, Florian Dörfler, Alessandro Chiuso, and Keyou You. Data-enabled policy optimization for direct adaptive learning of the lqr. *arXiv preprint arXiv:2401.14871*, 2024. URL https://arxiv.org/abs/2401.14871.