Statistically Consistent Inverse Optimal Control for Linear-Quadratic Tracking with Random Time Horizon

Han Zhang[1], Axel Ringh[2], Weihan Jiang[1], Shaoyuan Li[1], Xiaoming Hu[3]

1. Department of Automation, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
E-mail: {zhanghan_tc, JiangWH9, syli}@sjtu.edu.cn

2. Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96, Gothenburg, Sweden
E-mail: axelri@chalmers.se

3. Department of Mathematics, KTH Royal Institute of Technology, SE-100 44, Stockholm, Sweden
E-mail: hu@kth.se

Abstract: The goal of Inverse Optimal Control (IOC) is to identify the underlying objective function based on observed optimal trajectories. It provides a powerful framework to model expert's behavior, and a data-driven way to design an objective function so that the induced optimal control is adapted to a contextual environment. In this paper, we design an IOC algorithm for linear-quadratic tracking problems with random time horizon, and prove the statistical consistency of the algorithm. More specifically, the proposed estimator is the solution to a convex optimization problem, which means that the estimator does not suffer from local minima. This enables the proven statistical consistency to actually be achieved in practice. The algorithm is also verified on simulated data as well as data from a real world experiment, both in the setting of identifying the objective function of human tracking locomotion. The statistical consistency is illustrated on the synthetic data set, and the experimental results on the real data shows that we can get a good prediction on human tracking locomotion based on estimating the objective function. It shows that the theory and the model have a good performance in real practice. Moreover, the identified model can be used as a control target in personalized rehabilitation robot controller design, since the identified objective function describes personal habit and preferences.

Key Words: Inverse optimal control, trajectory tracking, system identification, convex optimization, semidefinite programming, human locomotion modelling, rehabilitation

1   Introduction

Optimal control is a powerful framework in which control actions are selected in order to minimize some given objective function; see, e.g., one of the monographs [1, 2]. In fact, many processes in nature have been observed to be optimal with respect to some criteria [3]. However, in engineering applications of optimal control, a *fundamental* problem is to design an appropriate objective function: it needs to be adapted to the contextual environment in which the system is operating in order to induce an appropriate control response. This is a difficult task, which relies heavily on the designers imagination. One way to overcome this would be to, instead of designing the cost criteria, identify it from observations of an expert system that behaves "optimally" in the environment. The latter is known as Inverse Optimal Control (IOC) [4], and has received considerable attention. In particular, the linear-quadratic problem has been studied in many different settings, including the infinite-horizon case in both continuous time [1, 5] and discrete time [6], respectively, as well as the finite-horizon case in both continuous time [7, 8] and discrete time [9–12], respectively. More general underlying dynamics and objective functions have also been considered [9, 13–17] and applied in areas such as, e.g., machine learning [18, 19], and to model and analyze human locomotion of different forms [20, 21].

However, all of the above IOC frameworks have limitations, in particular if one wants to apply IOC in a real-world situation. More precisely, any real-world data would inevitably contain noise: it can be process noise, observation noise, or both. Therefore, from a robustness and accuracy perspective, it is important to have an unbiased estimator which is statistically consistent, i.e., that converges to the true underlying parameter values when the number of observation increases. To the best of our knowledge, none of the IOC frameworks with more general underlying dynamics and objective functions have considered this aspect. Furthermore, they all suffer from the fact that the estimation problems are nonconvex, and hence a globally optimal solution cannot be guaranteed in practice. Regarding the literature on linear-quadratic IOC, most of them consider the stabilization problem. However, in many experimental set-ups, it is of greater interest to have the subject under investigation track a time-varying reference signal and to identify the corresponding objective function. Finally, real-world data can be of different time lengths, and this needs to be handled in a systematic way in order not to deteriorate the estimates.

In this work, we address these issues. More specifically, we consider the case of linear-quadratic, discrete-time IOC. The observations are obtained from expert systems tracking a given reference signal, which naturally also puts us in a finite-horizon setting since the reference signal is of finite length. Moreover, we model the data as arising from a system with process noise. Finally, in order to handle data of different time lengths, we also model the planing horizon in the forward problem as stochastic. Despite this, we can formulate the estimator as the solution to a convex optimization problem, and we show that it is statistically consistent. Lastly, we demonstrate the novelty and effectiveness of the method on a real-world experiment, where we identify the

objective function of a person performing rotational motions of the elbow to track a given reference signal.

Notations: $\mathbb{Z}_+$ denotes positive integers, and $\mathbb{S}_+^n$ denote the sets of $n \times n$ positive semi-definite matrices. We use $\succeq$ to denote the Loewner partial order on $\mathbb{S}_+^n$, i.e., for $G_1, G_2 \in \mathbb{S}^n$, $G_1 \succeq G_2$ means that $G_1 - G_2 \in \mathbb{S}_+^n$, and $G_1 \succ G_2$ means that $G_1 - G_2$ is strictly positive definite. $\|\cdot\|$ denotes the l_2-norm and $\|\cdot\|_F$ denotes the Frobenius norm. $\mathbf{1}_m$ denotes an all-one vector of dimension m, and $\otimes$ denotes Kronecker product. Further, we use ***italic bold font*** to denote stochastic vectors, and we use $\mathbb{E}_{\boldsymbol{x}}[\cdot]$ and $\mathbb{E}_{\boldsymbol{x}|\boldsymbol{y}}[\cdot]$ to denote the expected value under the distribution $\mathbb{P}(\boldsymbol{x})$ and the conditional distribution $\mathbb{P}(\boldsymbol{x}|\boldsymbol{y})$, respectively. The conditional covariance is denoted as $\mathrm{cov}_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z}}(\boldsymbol{x},\boldsymbol{y}) := \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z}}[(\boldsymbol{x} - \mathbb{E}_{\boldsymbol{x}|\boldsymbol{z}}[\boldsymbol{x}])(\boldsymbol{y} - \mathbb{E}_{\boldsymbol{y}|\boldsymbol{z}}[\boldsymbol{y}])^T]$. Finally, $\xrightarrow{p}$ denotes convergence in probability.

2 Problem formulation

We start by introducing the mathematical formulation of the forward optimal control problem. To this end, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space that carries random vectors $\bar{\boldsymbol{x}} \in \mathbb{R}^n$, $\{\boldsymbol{w}_t \in \mathbb{R}^n\}_{t=1}^{\infty}$, and a random variable $\mathcal{N} \in \{\nu_1, \nu_1 + 1, \cdots, \nu_2\} \subset \mathbb{Z}_+$. For each realization $(\bar{x}, N)$ of the random element $(\bar{\boldsymbol{x}}, \mathcal{N})$, corresponding to the initial position and planning horizon length, suppose that the tracking control of an agent is determined by a stochastic linear-quadratic control problem with respect to some known reference signal $\{x_t^r\}_{t=1}^{\nu_2}$, namely,

$$\min_{\substack{\boldsymbol{x}_{1:\nu_2}, \\ \boldsymbol{u}_{1:\nu_2}}} \quad \mathbb{E}_{\boldsymbol{w}_{\nu_2-N+1:\nu_2-1}} \left[\frac{1}{2}(\boldsymbol{x}_{\nu_2} - x_{\nu_2}^r)^T \bar{Q}(\boldsymbol{x}_{\nu_2} - x_{\nu_2}^r) \right. \tag{1a}$$

$$\left. + \sum_{t=\nu_2-N+1}^{\nu_2-1} \left[\frac{1}{2}(\boldsymbol{x}_t - x_t^r)^T \bar{Q}(\boldsymbol{x}_t - x_t^r) + \frac{1}{2}\|\boldsymbol{u}_t\|^2 \right] \right]$$

$$\text{s.t.} \quad \boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B(\boldsymbol{u}_t + \boldsymbol{w}_t), \ t = \nu_2 - N + 1 : \nu_2 - 1, \tag{1b}$$

$$\boldsymbol{x}_{\nu_2-N+1} = \bar{x}, \tag{1c}$$

$$\boldsymbol{x}_1 = \ldots = \boldsymbol{x}_{\nu_2-N} = 0, \tag{1d}$$

$$\boldsymbol{u}_1 = \ldots = \boldsymbol{u}_{\nu_2-N} = 0. \tag{1e}$$

Here, $\bar{Q} \in \mathbb{S}_+^n$, and $\boldsymbol{u}_t \in \mathbb{R}^m$ is the control signal. Notably, (1) means that the agent starts tracking the reference signal $\{x_t^r\}_{t=1}^{\nu_2}$ at the time instant $\nu_2 - \mathcal{N} + 1$, which is stochastic, and hence that the tracking is done over different planning horizons but with *the same terminal cost* and *running cost*. More precisely, given the time-horizon length $\mathcal{N} = N$ and initial value $\bar{\boldsymbol{x}} = \bar{x}$, the distribution of the agent's optimal trajectory $\mathbb{P}(\boldsymbol{x}_{1:N}^{\star} \mid \mathcal{N} = N, \bar{\boldsymbol{x}} = \bar{x})$ and optimal control $\mathbb{P}(\boldsymbol{u}_{1:N-1}^{\star} \mid \mathcal{N} = N, \bar{\boldsymbol{x}} = \bar{x})$ are implicitly given by solving (1), where we for convenience set $\boldsymbol{x}_t = 0$ and $\boldsymbol{u}_t = 0$ for time points t before the tracking starts (see (1d) and (1e)). In this context, we make the following mild assumptions on the stochastic planning horizon.

Assumption 1 (Stochastic planning horizon)**.** *The constant $\nu_2 \in \mathbb{Z}_+$ is known, and $\nu_2 \geq n + 1$. Moreover, the probability distribution for $\mathcal{N}$ satisfies $\mathbb{P}(\mathcal{N} \in [\nu_1, \nu_2]) = 1$, and $\mathbb{P}(\mathcal{N} = \nu_2) > 0$, where $\nu_1 \in \mathbb{Z}_+$ and $\nu_1 \leq \nu_2$.*

The above assumptions means that the longest possible planning horizon is known, that it is sufficiently long, and

that it can actually be realized, i.e., that it has a nonzero probability. Moreover, under mild regularity conditions on the probability distribution of $(\bar{\boldsymbol{x}}, \mathcal{N})$, the formulation in (1) defines joint probability distributions for $(\boldsymbol{x}_{1:\mathcal{N}}^{\star}, \mathcal{N}, \bar{\boldsymbol{x}})$ and $(\boldsymbol{u}_{1:\mathcal{N}-1}^{\star}, \mathcal{N}, \bar{\boldsymbol{x}})$ (cf. [22, Thm. 5.3]). From now on, in order to ease the notation, we omit the "star" in the superscript in the agents optimal states and control inputs.

Next, it is well-known that a linear-quadratic tracking problem can be rewritten as a linear-quadratic problem with a time-varying linear cost term. To this end, by introducing $\bar{q}_t := -\bar{Q}x_t^r$, for $t = 1, \ldots, \nu_2$, and omitting the constant terms, the objective function (1a) can be re-written as

$$\mathbb{E}_{\boldsymbol{w}_{\nu_2-N+1:\nu_2-1}} \left[\frac{1}{2}\boldsymbol{x}_{\nu_2}^T \bar{Q} \boldsymbol{x}_{\nu_2} + \bar{q}_{\nu_2}^T \boldsymbol{x}_{\nu_2} \right.$$

$$\left. + \sum_{t=\nu_2-N+1}^{\nu_2-1} \frac{1}{2}\boldsymbol{x}_t \bar{Q} \boldsymbol{x}_t + \bar{q}_t^T \boldsymbol{x}_t + \frac{1}{2}\|\boldsymbol{u}_t\|^2 \right]. \tag{2}$$

The IOC problem we consider in this work is: *given observations of trajectories generated as solutions to the forward problem* (1)*, form a statistically consistent estimate of the parameter $\bar{Q}$ in the objective function.* In order for this to be a well-posed problem, we need a number of assumptions.

Assumption 2 (System dynamics)**.** *The system (A, B) is controllable, B has full column rank, and A is invertible.*

Controllability is a standard assumption, and the fact that B has full column rank means that there are no redundant control signals (in particular, the system is not over-actuated). The last assumption is motivated by the fact that a discrete-time system is often sampled from a continuous-time system, and the discretized A matrix is always invertible. In addition, we also make the following assumptions regarding the random elements and the parameter that we want to estimate.

Assumption 3 (I.I.D random variables)**.** *The random elements $(\bar{\boldsymbol{x}}, \mathcal{N})$ and $\{\boldsymbol{w}_t\}_{t=1}^{\infty}$ are all independent. In addition, the random vectors $\{\boldsymbol{w}_t\}_{t=1}^{\infty}$ are identically distributed (I.I.D), $\mathbb{E}[\boldsymbol{w}_t] = 0$, $\forall t$, and $\mathrm{cov}(\boldsymbol{w}_t, \boldsymbol{w}_t) = \Sigma_w \succ 0$, where Σ_w is a priori known and $\|\Sigma_w\|_F < \infty$.*

Assumption 4 (Persistent excitation)**.** *It holds that $\mathrm{cov}_{\bar{\boldsymbol{x}}|\mathcal{N}=\nu_2}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{x}}) \succ 0$ and $\mathbb{E}[\|\bar{\boldsymbol{x}}\|^2] < \infty$.*

Assumption 5 (Bounded parameter)**.** *The unknown "true" parameter $\bar{Q}$ that governs the agents tracking behavior lives in the compact set*

$$\mathbb{G}(\varphi) := \left\{ \bar{Q} \in \mathbb{S}_+^n \mid \|\bar{Q}\|_F \leq \varphi \right\},$$

for some (potentially unknown) $0 < \varphi < \infty$.

3 Main results

The optimal control signal that solves (1) turns out to have the same form as in the deterministic case, i.e., without the process noise $\boldsymbol{w}_t$ in the dynamics (1b). This can be seen by following a derivation similar to the one in [2, Sec. 4.1]. In particular, conditioned on $\mathcal{N} = \nu_2$ we have that for $t = 1 : \nu_2 - 1$, the optimal control signal is given by

$$\boldsymbol{u}_t = -(B^T \bar{P}_{t+1} B + I)^{-1} \left(B^T \bar{P}_{t+1} A \boldsymbol{x}_t + B^T \bar{\eta}_{t+1} \right), \tag{3}$$

where the sequence of $\bar{P}_{1:N}$ and $\bar{\eta}_{1:N}$ are generated by the Riccati iterations

$$\bar{P}_N = \bar{Q}, \tag{4a}$$

$$\begin{aligned}\bar{P}_t = &\, A^T \bar{P}_{t+1} A + \bar{Q} \\ &- A^T \bar{P}_{t+1} B (B^T \bar{P}_{t+1} B + I)^{-1} B^T \bar{P}_{t+1} A, \\ &t = 1 : N - 1;\end{aligned} \tag{4b}$$

$$\bar{\eta}_N = \bar{q}_N, \tag{4c}$$

$$\begin{aligned}\bar{\eta}_t = &\, \left(A - B (B^T \bar{P}_{t+1} B + I)^{-1} B^T \bar{P}_{t+1} A \right)^T \bar{\eta}_{t+1} \\ &+ \bar{q}_t, \quad t = 1 : N - 1.\end{aligned} \tag{4d}$$

In view of the special form of the linear-quadratic optimal control, we propose to solve the following convex optimization problem to reconstruct the "true" parameter $\bar{Q}$:

$$\min_{\substack{Q \in \mathbb{G}(\varphi) \\ \{q_t \in \mathbb{R}^n\}_{t=1:\nu_2} \\ \{P_t \in \mathbb{S}_+^n\}_{t=1:\nu_2}, \\ \{\eta_t \in \mathbb{R}^n\}_{t=1:\nu_2}, \\ \{\xi_t \in \mathbb{R}\}_{t=1:\nu_2-1}}} \Psi(Q, q_{1:\nu_2}, P_{1:\nu_2}, \eta_{1:\nu_2}, \xi_{1:\nu_2-1})$$

$$\text{s.t.} \quad P_{\nu_2} = Q, \quad \eta_{\nu_2} = q_{\nu_2}, \tag{5a}$$

$$q_t = -Q x_t^T, \quad t = 1 : \nu_2 \tag{5b}$$

$$H_t := \begin{bmatrix} \mathfrak{R}_t & \mathfrak{S}_t & g_t \\ \mathfrak{S}_t^T & A^T P_{t+1} A + Q - P_t & \beta_t \\ g_t^T & \beta_t^T & \xi_t \end{bmatrix} \succeq 0,$$

$$t = 1 : \nu_2 - 1. \tag{5c}$$

where $\mathfrak{R}_t := B^T P_{t+1} B + I$, $\mathfrak{S}_t := B^T P_{t+1} A$, $g_t := B^T \eta_{t+1}$, $\beta_t := q_t + A^T \eta_{t+1} - \eta_t$, and

$$\begin{aligned}\Psi(Q, q_{1:\nu_2}, P_{1:\nu_2}, \eta_{1:\nu_2}, \xi_{1:\nu_2-1}) := \, &\mathbb{E}_{\boldsymbol{x}_{1:\nu_2}, \mathcal{N}} \Big[\frac{1}{2} \boldsymbol{x}_{\nu_2}^T P_{\nu_2} \boldsymbol{x}_{\nu_2} \\ + \eta_{\nu_2}^T \boldsymbol{x}_{\nu_2} - \frac{1}{2} \boldsymbol{x}_{\nu_2-\mathcal{N}+1}^T P_{\nu_2-\mathcal{N}+1} \boldsymbol{x}_{\nu_2-\mathcal{N}+1} \\ - \eta_{\nu_2-\mathcal{N}+1}^T \boldsymbol{x}_{\nu_2-\mathcal{N}+1} + \sum_{t=\nu_2-\mathcal{N}+1}^{\nu_2-1} \Big(\frac{1}{2} \xi_t + \frac{1}{2} \boldsymbol{x}_t^T Q \boldsymbol{x}_t \\ + q_t^T \boldsymbol{x}_t - \frac{1}{2} \operatorname{tr}(B^T P_{t+1} B \Sigma_w) \Big) \Big].\end{aligned} \tag{5d}$$

Before we proceed, we have the following intermediate result that will be useful in the analysis.

Lemma 1. *Let $\tilde{\boldsymbol{x}}_t := [\boldsymbol{x}_t^T, 1]^T$ and $\bar{\tilde{\boldsymbol{x}}}_{t+1} := [(A\boldsymbol{x}_t + B\boldsymbol{u}_t)^T, 1]^T$. Under Assumptions 1 and 4, it holds that $\mathbb{E}_{\boldsymbol{x}_t | \mathcal{N} = \nu_2}[\tilde{\boldsymbol{x}}_t \tilde{\boldsymbol{x}}_t^T] \succ 0$, $\mathbb{E}_{\boldsymbol{x}_t | \mathcal{N} = \nu_2}[\bar{\tilde{\boldsymbol{x}}}_{t+1} \bar{\tilde{\boldsymbol{x}}}_{t+1}^T] \succ 0$, and $\mathbb{E}[\|\tilde{\boldsymbol{x}}_t\|^2] < \infty$ for all $t = 1 : \nu_2$.*

Proof. Due to space limitation, the proof is omitted. $\square$

Theorem 1. *Let $(\bar{Q}, \bar{q}_{1:\nu_2})$ be the "true" parameters of the stochastic linear-quadratic regulator that governs the agent, and let $\bar{P}_{1:\nu_2}$, $\bar{\eta}_{1:\nu_2}$ be generated by the corresponding Riccati iterations (4). Under Assumptions 1, 2, 3, 4 and 5, $(\bar{Q}, \bar{q}_{1:\nu_2}, \bar{P}_{1:\nu_2}, \bar{\eta}_{1:\nu_2})$ is the unique optimal solution to (5).*

Proof. To ease the notation, in the proof we omit the arguments of the objective function and simply write $\Psi(\cdot)$. Moreover, to prove the statement of the theorem we show i) that $\Psi(\cdot)$ is bounded from below on the feasible region, ii) that the true parameter values attain this bound, and iii) that the solution is unique.

To this end, first note that (5d) can be written as

$$\Psi(\cdot) = \sum_{N=\nu_1}^{\nu_2} \mathbb{P}(\mathcal{N} = N) \, \mathbb{E}_{\boldsymbol{x}_{\nu_2-N+1:\nu_2} | \mathcal{N} = N} [\psi_N(\cdot)], \tag{6}$$

where $\psi_N(\cdot) = \sum_{t=\nu_2-N+1}^{\nu_2-1} \psi_{t,N}(\cdot)$ and

$$\begin{aligned}\psi_{t,N}(\cdot) = &\, \frac{1}{2} \boldsymbol{x}_{t+1}^T P_{t+1} \boldsymbol{x}_{t+1} + \eta_{t+1}^T \boldsymbol{x}_{t+1} - \frac{1}{2} \boldsymbol{x}_t^T P_t \boldsymbol{x}_t - \eta_t^T \boldsymbol{x}_t \\ &+ \frac{1}{2} \xi_t + \frac{1}{2} \boldsymbol{x}_t^T Q \boldsymbol{x}_t + q_t^T \boldsymbol{x}_t - \frac{1}{2} \operatorname{tr}(B^T P_{t+1} B \Sigma_w),\end{aligned}$$

and where the expectation in (6) is only taken over time points $\nu_2 - N + 1 : \nu_2$ since the probability measure of $\mathbb{P}(\boldsymbol{x}_{1:\nu_2-N} | \mathcal{N} = N)$ is marginalized out. In view of the expression of $\psi_{t,N}(\cdot)$ and (1b), we can further write (6) as

$$\begin{aligned}\Psi(\cdot) &= \sum_{N=\nu_1}^{\nu_2} \mathbb{P}(\mathcal{N} = N) \sum_{t=\nu_2-N+1}^{\nu_2-1} \mathbb{E}_{\boldsymbol{x}_{t:t+1} | \mathcal{N} = N} [\psi_{t,N}(\cdot)] \\ &= \sum_{N=\nu_1}^{\nu_2} \mathbb{P}(\mathcal{N} = N) \sum_{t=\nu_2-N+1}^{\nu_2-1} \mathbb{E}_{\boldsymbol{w}_t, \boldsymbol{x}_t | \mathcal{N} = N} [\psi_{t,N}(\cdot)] \tag{7}\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}_{\boldsymbol{w}_t, \boldsymbol{x}_t | \mathcal{N} = N} [\psi_{t,N}(\cdot)] = &\, \mathbb{E}_{\boldsymbol{x}_t | \mathcal{N} = N} \Big[\mathbb{E}_{\boldsymbol{w}_t} \Big[\frac{1}{2} \left(A\boldsymbol{x}_t + B(\boldsymbol{u}_t + \boldsymbol{w}_t) \right)^T \\ &\times P_{t+1} \left(A\boldsymbol{x}_t + B(\boldsymbol{u}_t + \boldsymbol{w}_t) \right) + \eta_{t+1}^T \left(A\boldsymbol{x}_t + B(\boldsymbol{u}_t + \boldsymbol{w}_t) \right) \\ &- \frac{1}{2} \boldsymbol{x}_t^T P_t \boldsymbol{x}_t - \eta_t^T \boldsymbol{x}_t + \frac{1}{2} \xi_t + \frac{1}{2} \boldsymbol{x}_t^T Q \boldsymbol{x}_t + q_t^T \boldsymbol{x}_t \\ &- \frac{1}{2} \operatorname{tr}(B^T P_{t+1} B \Sigma_w) \Big] \Big],\end{aligned}$$

since by Assumption 3 the noise $\{\boldsymbol{w}_t\}_{t=1}^{\infty}$ is independent of any other stochastic element. Now, using Assumption 3 and the cyclic permutation property of the matrix trace operator, we know that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{w}_t}[\boldsymbol{w}_t^T B^T P_{t+1} B \boldsymbol{w}_t] &= \mathbb{E}_{\boldsymbol{w}_t}[\operatorname{tr}(\boldsymbol{w}_t^T B^T P_{t+1} B \boldsymbol{w}_t)] \\ &= \mathbb{E}_{\boldsymbol{w}_t}[\operatorname{tr}(B^T P_{t+1} B \boldsymbol{w}_t \boldsymbol{w}_t^T)] = \operatorname{tr}(B^T P_{t+1} B \Sigma_w)\end{aligned}$$

which together with $\mathbb{E}_{\boldsymbol{w}_t}[\boldsymbol{w}_t] = 0$ implies that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{w}_t, \boldsymbol{x}_t | \mathcal{N} = N} [\psi_{t,N}(\cdot)] = &\, \mathbb{E}_{\boldsymbol{x}_t | \mathcal{N} = N} \Big[\frac{1}{2} \left(A\boldsymbol{x}_t + B\boldsymbol{u}_t \right)^T \\ &\times P_{t+1} \left(A\boldsymbol{x}_t + B\boldsymbol{u}_t \right) + \eta_{t+1}^T \left(A\boldsymbol{x}_t + B\boldsymbol{u}_t \right) \\ &- \frac{1}{2} \boldsymbol{x}_t^T P_t \boldsymbol{x}_t - \eta_t^T \boldsymbol{x}_t + \frac{1}{2} \xi_t + \frac{1}{2} \boldsymbol{x}_t^T Q \boldsymbol{x}_t + q_t^T \boldsymbol{x}_t \Big] \\ = &\, \mathbb{E}_{\boldsymbol{x}_t | \mathcal{N} = N} \Big[\frac{1}{2} \begin{bmatrix} \boldsymbol{u}_t^T & \boldsymbol{x}_t^T & 1 \end{bmatrix} H_t \begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{x}_t \\ 1 \end{bmatrix} - \frac{1}{2} \|\boldsymbol{u}_t\|^2 \Big],\end{aligned}$$

where H_t is the matrix given in (5c). Next, note that for any feasible P_t it holds that $P_t \in \mathbb{S}_+^n$, and hence $\mathfrak{R}_t = B^T P_{t+1} B + I \succ 0$ which means that it is invertible. In particular, this means that we can take the Schur complement of H_t with respect to $\mathfrak{R}_t$, which gives

$$H_t \backslash \mathfrak{R}_t = \tag{8}$$
$$\begin{bmatrix} A^T P_{t+1} A + Q - P_t & \beta_t \\ \beta_t^T & \xi_t \end{bmatrix} - \begin{bmatrix} \mathfrak{S}_t^T \\ g_t^T \end{bmatrix} \mathfrak{R}_t^{-1} \begin{bmatrix} \mathfrak{S}_t & g_t \end{bmatrix}.$$

The above expression can therefore be further rewritten as

$$\underset{\boldsymbol{w}_t,\boldsymbol{x}_t|\mathcal{N}=N}{\mathbb{E}}[\psi_{t,N}(\cdot)] = \underset{\boldsymbol{x}_t|\mathcal{N}=N}{\mathbb{E}}\left[\frac{1}{2}\begin{bmatrix}\boldsymbol{u}_t^T & \boldsymbol{x}_t^T & 1\end{bmatrix} G_t\times\right.$$

$$\left.\begin{bmatrix}\mathfrak{R}_t & \\ & H_t\backslash\mathfrak{R}_t\end{bmatrix}G_t^T\begin{bmatrix}\boldsymbol{u}_t\\\boldsymbol{x}_t\\1\end{bmatrix}-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right], \quad (9)$$

where

$$G_t = \begin{bmatrix}I & & \\ \mathfrak{S}_t^T\mathfrak{R}_t^{-1} & I & \\ g_t^T\mathfrak{R}_t^{-1} & & 1\end{bmatrix}.$$

Now, using the notation $\bar{\bar{\boldsymbol{x}}}_{t+1}$ (introduced in Lemma 1), we can expand (9) and therefore rewrite (7) as

$$\Psi(\cdot) = \sum_{N=\nu_1}^{\nu_2}\left\{\mathbb{P}(\mathcal{N}=N)\sum_{t=\nu_2-N+1}^{\nu_2-1}\underset{\boldsymbol{x}_t|\mathcal{N}=N}{\mathbb{E}}\left[-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right.\right.$$

$$\left.\left.+\frac{1}{2}\bar{\bar{\boldsymbol{x}}}_t^T(H_t\backslash\mathfrak{R}_t)\bar{\bar{\boldsymbol{x}}}_t+\frac{1}{2}\|\mathfrak{R}_t^{1/2}(\boldsymbol{u}_t+\mathfrak{R}_t^{-1}\mathfrak{S}_t\boldsymbol{x}_t+\mathfrak{R}_t^{-1}g_t)\|^2\right.\right].$$
$$(10)$$

Finally, note that since $\mathfrak{R}_t \succ 0$, the Schur complement $H_t\backslash\mathfrak{R}_t \succeq 0$ [23, p. 495], and by also using (1e) it hence follows that

$$\Psi(\cdot) \geq \sum_{N=\nu_1}^{\nu_2}\left\{\mathbb{P}(\mathcal{N}=N)\sum_{t=1}^{\nu_2-1}\underset{\boldsymbol{x}_t|\mathcal{N}=N}{\mathbb{E}}\left[-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right]\right\}$$

$$= \sum_{t=1}^{\nu_2-1}\underset{\boldsymbol{x}_t}{\mathbb{E}}\left[-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right], \quad (11)$$

where the inequality follows by simply adding the zeros $\sum_{t=1}^{\nu_2-N}\mathbb{E}_{\boldsymbol{x}_t|\mathcal{N}=N}\left[-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right]$ and removing all other non-negative terms. This gives a lower bound for the optimal value.

Next, we show that the lower bound is actually attained by the "true" $(\bar{Q}, \bar{q}_{1:\nu_2}, \bar{P}_{1:\nu_2}, \bar{\eta}_{1:\nu_2})$. First note that the Riccati iteration (4) with the "true" $\bar{Q}$ and $\bar{q}$ can be written as

$$\begin{bmatrix}\bar{P}_N & \bar{\eta}_N\\\bar{\eta}_N^T & \bar{\xi}_N\end{bmatrix} = \begin{bmatrix}\bar{Q} & \bar{q}\\\bar{q}^T & 0\end{bmatrix}, \quad (12a)$$

$$\boldsymbol{0} = \begin{bmatrix}A^T\bar{P}_{t+1}A + \bar{Q} - \bar{P}_t & \bar{q}+A^T\bar{\eta}_{t+1}-\bar{\eta}_t\\q^T+\bar{\eta}_{t+1}^TA-\bar{\eta}_t^T & \bar{\xi}_t\end{bmatrix} \quad (12b)$$

$$-\begin{bmatrix}A^T\bar{P}_{t+1}B\\\bar{\eta}_{t+1}^TB\end{bmatrix}(B^T\bar{P}_{t+1}B+I)^{-1}\begin{bmatrix}B^T\bar{P}_{t+1}A & B^T\bar{\eta}_{t+1}\end{bmatrix},$$

with a corresponding appropriate iteration for $\bar{\xi}_t$. Moreover, as already observed earlier, it holds that $\bar{\mathfrak{R}}_t = B^T\bar{P}_{t+1}B + I \succ 0$ for all $t = 1 : N-1$ since $\bar{P}_t \in \mathbb{S}_+^n$. Now, comparing (8) and (12b) we see that the latter is the Schur complement of $\bar{H}_t$ with respect to the top left corner, i.e., $\bar{H}_t\backslash\bar{\mathfrak{R}}_t$, and hence we have that $\bar{H}_t\backslash\bar{\mathfrak{R}}_t = 0 \succeq 0$. Therefore, by [23, p. 495], this implies that $\bar{H}_t \succeq 0, \forall t = 1 : N-1$. Together with (12a) this implies that $\bar{Q}$, $\bar{q}$ and $\bar{P}_{1:N}$, $\bar{\eta}_{1:N}$ generated by Riccati iteration (4) is a feasible solution to (5). Further, in view of (10) and the form of optimal control (3), we can conclude that the lower bound $\sum_{t=1}^{\nu_2-1}\mathbb{E}_{\boldsymbol{x}_t}\left[-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right]$ is attained by the "true" $(\bar{Q}, \bar{q}_{1:\nu_2}, \bar{P}_{1:\nu_2}, \bar{\eta}_{1:\nu_2})$.

Finally, we show the uniqueness. To this end, suppose there exists another optimal solution $(Q' = \bar{Q} + \Delta Q, \{q_t' =$

$\bar{q}_t - \Delta Q x_t^r\}_{t=1}^{\nu_2}, \{P_t' = \bar{P}_t + \Delta P_t\}_{t=1}^{\nu_2}, \{\eta_t' = \bar{\eta}_t + \Delta\eta_t\}_{t=1}^{\nu_2})$ that attains the same optimal value. In particular, this point must also attain the lower bound in (11). Moreover, by repeating the arguments for the inequality in (11), but removing all non-negative terms except for $\mathcal{N} = \nu_2$, we get

$$\sum_{t=1}^{\nu_2-1}\underset{\boldsymbol{x}_t}{\mathbb{E}}\left[-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right] = \Psi(Q', q', P_{1:\nu_2}', \eta_{1:\nu_2}')$$

$$\geq \sum_{t=1}^{\nu_2-1}\underset{\boldsymbol{x}_t}{\mathbb{E}}\left[-\frac{1}{2}\|\boldsymbol{u}_t\|^2\right] + \mathbb{P}(\mathcal{N}=\nu_2)\sum_{t=1}^{\nu_2-1}\underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\right.$$

$$\left.\frac{1}{2}\bar{\bar{\boldsymbol{x}}}_t^T(H_t'\backslash\mathfrak{R}_t')\bar{\bar{\boldsymbol{x}}}_t+\frac{1}{2}\|(\mathfrak{R}_t')^{1/2}(\boldsymbol{u}_t+\mathfrak{R}_t'^{-1}\mathfrak{S}_t'\boldsymbol{x}_t+\mathfrak{R}_t'^{-1}g_t')\|^2\right].$$

Since $H_t'\backslash\mathfrak{R}_t' \succeq 0, \forall t = 1 : \nu_2$, all terms in the last sum are nonnegative. Furthermore, since by assumption $\mathbb{P}(\mathcal{N}=\nu_2) > 0$, it must hold that all terms in the last sum are equal to zero. More precisely, we must have that

$$\underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\bar{\bar{\boldsymbol{x}}}_t^T(H_t'\backslash\mathfrak{R}_t')\bar{\bar{\boldsymbol{x}}}_t\right] = \underset{\boldsymbol{x}_t|\mathcal{N}=N}{\mathbb{E}}\left[\text{tr}((H_t'\backslash\mathfrak{R}_t')\bar{\bar{\boldsymbol{x}}}_t\bar{\bar{\boldsymbol{x}}}_t^T)\right]$$

$$= \text{tr}\left((H_t'\backslash\mathfrak{R}_t')\underset{\boldsymbol{x}_t|\mathcal{N}=N}{\mathbb{E}}\left[\bar{\bar{\boldsymbol{x}}}_t\bar{\bar{\boldsymbol{x}}}_t^T\right]\right) = 0 \quad (13a)$$

$$\underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\|(\mathfrak{R}_t')^{1/2}(\boldsymbol{u}_t+\mathfrak{R}_t'^{-1}\mathfrak{S}_t'\boldsymbol{x}_t+\mathfrak{R}_t'^{-1}g_t')\|^2\right] = 0.$$
$$(13b)$$

Since by Lemma 1, $\mathbb{E}_{\boldsymbol{x}_t|\mathcal{N}=\nu_2}[\bar{\bar{\boldsymbol{x}}}_t\bar{\bar{\boldsymbol{x}}}_t^T] \succ 0$, in order for (13a) to hold, we must have that $H_t'\backslash\mathfrak{R}_t' = 0, \forall t = 1 : \nu_2$, which implies that the Riccati iterations (4) must hold for all optimal solutions to (5) (cf. (12b)).

To ease notation when analyzing expression (13b), we temporarily introduce $\mathfrak{a}_t := \boldsymbol{u}_t+\mathfrak{R}_t'^{-1}\mathfrak{S}_t'\boldsymbol{x}_t+\mathfrak{R}_t'^{-1}g_t'$. From (13b) it then follows that

$$0 = \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\|(\mathfrak{R}_t')^{1/2}\mathfrak{a}_t\|^2\right] = \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\mathfrak{a}_t^T\mathfrak{R}_t'\mathfrak{a}_t\right]$$

$$= \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\text{tr}(\mathfrak{a}_t^T\mathfrak{R}_t'\mathfrak{a}_t)\right] = \text{tr}(\mathfrak{R}_t'\underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\mathfrak{a}_t\mathfrak{a}_t^T\right]).$$

Since $\mathfrak{R}_t' \succ 0$ and $\mathbb{E}_{\boldsymbol{x}_t|\mathcal{N}=\nu_2}[\mathfrak{a}_t\mathfrak{a}_t^T] \succeq 0$, we must have that $\mathbb{E}_{\boldsymbol{x}_t|\mathcal{N}=\nu_2}[\mathfrak{a}_t\mathfrak{a}_t^T] = \boldsymbol{0}$. This means that

$$0 = \text{tr}(\mathfrak{R}_t'\boldsymbol{0}\mathfrak{R}_t'^T) = \text{tr}\left(\mathfrak{R}_t'\underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\mathfrak{a}_t\mathfrak{a}_t^T\right]\mathfrak{R}_t'^T\right)$$

$$= \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\text{tr}(\mathfrak{R}_t'\mathfrak{a}_t\mathfrak{a}_t^T\mathfrak{R}_t'^T)\right] = \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\|\mathfrak{R}_t'\mathfrak{a}_t\|^2\right]$$

$$= \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\|\mathfrak{R}_t'\boldsymbol{u}_t+\mathfrak{S}_t'\boldsymbol{x}_t+g_t'\|^2\right]$$

$$= \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\|\boldsymbol{u}_t+B^TP_{t+1}'(A\boldsymbol{x}_t+B\boldsymbol{u}_t)+B^T\eta_{t+1}'\|^2\right],$$

where the last equality comes from simply using the expressions for $\mathfrak{R}_t'$, $\mathfrak{S}_t'$, and g_t', as well as grouping the terms. Moreover, by (3) we have that when conditioned on that $\mathcal{N} = \nu_2$, $\boldsymbol{u}_t = -B^T\bar{P}_{t+1}(A\boldsymbol{x}_t+B\boldsymbol{u}_t)-B^T\bar{\eta}_{t+1}$ for $t = 1 : \nu_2 - 1$. Plugging this into the above gives that

$$0 = \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\|B^T\Delta P_{t+1}(A\boldsymbol{x}_t+B\boldsymbol{u}_t)+B^T\Delta\eta_{t+1}\|^2\right]$$

$$= \underset{\boldsymbol{x}_t|\mathcal{N}=\nu_2}{\mathbb{E}}\left[\|B^T\begin{bmatrix}\Delta P_{t+1} & \Delta\eta_{t+1}\end{bmatrix}\bar{\bar{\boldsymbol{x}}}_{t+1}\|^2\right],$$

and rewriting this using the trace, as in (13a), as well as the fact that by Lemma 1 we have that $\mathbb{E}_{\boldsymbol{x}_t|\mathcal{N}=\nu_2}[\bar{\bar{\boldsymbol{x}}}_{t+1}\bar{\bar{\boldsymbol{x}}}_{t+1}^T] \succ 0$, we can conclude that

$$B^T \begin{bmatrix} \Delta P_{t+1} & \Delta\eta_{t+1} \end{bmatrix} = 0.$$

Now, the uniqueness can be proved by following the same argument as in [10, Thm. 2.1]. $\qquad\square$

Note that in reality, the distributions of $\bar{\boldsymbol{x}}$, $\boldsymbol{w}_t$ and $\mathcal{N}$ are usually not a priori known, and hence it is not possible to calculate the objective function (5d) explicitly. Therefore, we cannot solve (5) directly, but instead need to empirically estimate the objective function. To this end, let M be the total number of trajectories observed, and let M_N denote the number of observations which has N time steps. Clearly, $\sum_{N=\nu_1}^{\nu_2} M_N = M$. First, we therefore approximate the expectation over $\mathcal{N}$ by using the empirical mean. Considering (6), this is the same as estimating the probabilities $\mathbb{P}(\mathcal{N}=N)$ using the empirical estimates M_N/M. Then, following along [12, Sec. 4.2], for each value of N we approximate the inner expectation in (6) with the empirical mean

$$\frac{1}{M_N}\sum_{i_N=1}^{M_N}\Big[\frac{1}{2}\boldsymbol{x}_{\nu_2}^{i_N T}P_{\nu_2}\boldsymbol{x}_{\nu_2}^{i_N} + \eta_{\nu_2}^T\boldsymbol{x}_{\nu_2}^{i_N} - \frac{1}{2}\boldsymbol{x}_{\nu_2-N+1}^{i_N T}P_{\nu_2-N+1}$$

$$\times \boldsymbol{x}_{\nu_2-N+1}^{i_N} - \eta_{\nu_2-N+1}^T\boldsymbol{x}_{\nu_2-N+1}^{i_N} + \sum_{t=\nu_2-N+1}^{\nu_2-1}\Big(\frac{1}{2}\xi_t$$

$$+ \frac{1}{2}\boldsymbol{x}_t^{i_N T}Q\boldsymbol{x}_t^{i_N} + q_t^T\boldsymbol{x}_t^{i_N} - \frac{1}{2}\operatorname{tr}(B^T P_{t+1}B\Sigma_w)\Big)\Big].$$

Using both these approximations, we formulate the estimation problem as

$$\min_{\substack{Q\in\mathbb{G}(\varphi)\\ \{q_t\in\mathbb{R}^n\}_{t=1:\nu_2}\\ \{P_t\in\mathbb{S}_+^n\}_{t=1:\nu_2},\\ \{\eta_t\in\mathbb{R}^n\}_{t=1:\nu_2},\\ \{\xi_t\in\mathbb{R}\}_{t=1:\nu_2-1}}} \Psi^o(Q,q_{1:\nu_2},P_{1:\nu_2},\eta_{1:\nu_2},\xi_{1:\nu_2-1})$$

$$\text{s.t. } (5a)–(5c) \text{ hold}, \tag{14}$$

where

$$\Psi^o(Q,q_{1:\nu_2},P_{1:\nu_2},\eta_{1:\nu_2},\xi_{1:\nu_2-1}) = \frac{1}{M}\sum_{N=\nu_1}^{\nu_2}\sum_{i_N=1}^{M_N}\Big[$$

$$\frac{1}{2}\boldsymbol{x}_{\nu_2}^{i_N T}P_{\nu_2}\boldsymbol{x}_{\nu_2}^{i_N} + \eta_{\nu_2}^T\boldsymbol{x}_{\nu_2}^{i_N} - \frac{1}{2}\boldsymbol{x}_{\nu_2-N+1}^{i_N T}P_{\nu_2-N+1}$$

$$\times \boldsymbol{x}_{\nu_2-N+1}^{i_N} - \eta_{\nu_2-N+1}^T\boldsymbol{x}_{\nu_2-N+1}^{i_N} + \sum_{t=\nu_2-N+1}^{\nu_2-1}\Big(\frac{1}{2}\xi_t$$

$$+ \frac{1}{2}\boldsymbol{x}_t^{i_N T}Q\boldsymbol{x}_t^{i_N} + q_t^T\boldsymbol{x}_t^{i_N} - \frac{1}{2}\operatorname{tr}(B^T P_{t+1}B\Sigma_w)\Big)\Big].$$

Next, we show that (14) is a statistically consistent estimator of the true parameter $\bar{Q}$. To this end, we first have the following two lemmas. For brevity, the details in the proofs are omitted.

Lemma 2. *Let $\mathcal{D}$ be the set*

$$\{Q\in\mathbb{G}(\varphi), \{q_t\in\mathbb{R}^n\}_{t=1:\nu_2}, \{P_t\in\mathbb{S}_+^n\}_{t=1:\nu_2},$$
$$\{\eta_t\in\mathbb{R}^n\}_{t=1:\nu_2}, \{\xi_t\in\mathbb{R}\}_{t=1:\nu_2-1} : (5a)–(5c) \text{ hold}\}.$$

Then, $\mathcal{D}$ is compact and for any observed data the objective function $\Psi^o(Q,q_{1:\nu_2},P_{1:\nu_2},\eta_{1:\nu_2},\xi_{1:\nu_2-1})$ is bounded on $\mathcal{D}$.

Proof. This follows by adapting the arguments in [12, Proof of Lem. 4.1]. $\qquad\square$

Lemma 3 (Uniform law of large numbers). *Under Assumptions 1, 2, 3, 4 and 5, the optimal value*

$$\sup |\Psi^o(Q,q_t,P_t,\eta_t,\xi_t) - \Psi(Q,q_t,P_t,\eta_t,\xi_t)|$$
$$s.t. \ (Q,q_t,P_t,\eta_t,\xi_t)\in\mathcal{D}$$

converges to 0 almost surely as $M\to\infty$.

Proof. This follows by bounding the argument inside the expectation in $\Psi(Q,q_{1:\nu_2},P_{1:\nu_2},\eta_{1:\nu_2},\xi_{1:\nu_2-1})$, given in (5d), from above by an integrable function of the random variables and which is independent of the parameters $(Q,q_{1:\nu_2},P_{1:\nu_2},\eta_{1:\nu_2},\xi_{1:\nu_2-1})$, and then applying [24, Thm. 2]. The former can be done via an argument similar to [12, Proof of Lem. 4.2], which relies on the bounds from Lemma 1. $\qquad\square$

Theorem 2 (Statistical consistency). *Under Assumptions 1, 2, 3, 4, and 5, the optimization problem (14) admits an optimal solution. For any such optimal solution $(Q_M^\star,q_{1:\nu_2}^\star,P_{1:\nu_2}^\star,\eta_{1:\nu_2}^\star,\xi_{1:\nu_2-1}^\star)$, we have that $Q_M^\star\xrightarrow{p}\bar{Q}$ as $M\to\infty$.*

Proof. It can be readily seen that (14) is convex, and hence any locally optimal solution is globally optimal. Moreover, by Lemma 2, the feasible domain of (14) is compact and Ψ^o is bounded on the feasible domain. Hence, (14) attains an optimal solution. The result now follows from [25, Thm. 5.7]. In particular, since convergence almost surely implies convergence in probability [22, Lem. 3.2], Lemma 3 implies that the first condition in the preceding theorem is satisfied. Moreover, the fact that $\mathcal{D}$ is compact (by Lemma 2) together with the fact that the optimal solution to (5) is unique (by Theorem 1) implies that the second condition of the preceding theorem holds. Hence the result follows. $\qquad\square$

We end this section with a few remarks.

Remark 1. *Most of the analysis still holds if the state terminal and running cost in the objective function of forward problem (1) is of the form $\frac{1}{2}\boldsymbol{x}_t^T\bar{Q}\boldsymbol{x}_t + \bar{q}^T\boldsymbol{x}_t$, i.e., with a constant unknown linear cost term $\bar{q}$ that does not relate to $\bar{Q}$ as in the tracking problem. Extension to recover the unknown constant parameters $(\bar{Q},\bar{q})$ is straight-froward.*

Remark 2. *The IOC algorithm (14) can easily be extended to the case where the observed data is obtained from experiments performed with a number of different reference signals for the tracking. The estimator is still formulated as the solution to a convex optimization problem akin to (14), however for each reference signal we get a set of LMIs (5a)–(5c) as constraints.*

4 Simulations and real-world experiments

In this section we demonstrate the developed method, both on synthetic data in a simulation study and on real measurement data from an experimental set-up. The simulation study uses the same underlying model as the experimental set-up,
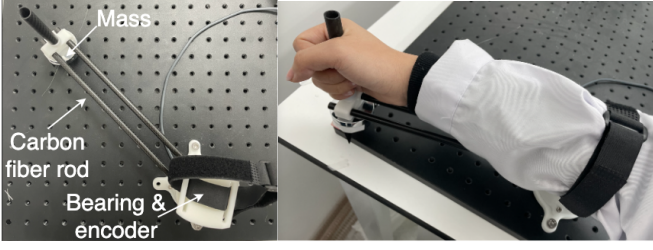
(a) The experimental device   (b) Human tracking locomotion

Fig. 1: (1a): A mass of 0.2kg is attached to a light carbon fiber rod and can rotate around the axis. The rotation is performed via a bearing to eliminate potential friction as much as possible. The distance between the centre of the mass and the axis is 0.255 metres. The angular position and velocity of the mass are measured every 0.05 seconds by an encoder attached to the axis. (1b): During the experiment, the test subject is using the device from Fig. 1a as shown in this image. Moreover, two sliding blocks that correspond to the reference signal and the state of the experimental device are shown on the computer screen. The human holds the handle and rotate his/her elbow to control the corresponding sliding block on the screen so that it tracks the reference signal.

and hence we first describe the latter. Since we do not know what the "true" $\bar{Q}$ is for a real agent in practice, we create synthetic data for simulation to illustrate the statistical consistency of proposed algorithm. The performance of the algorithm in real practice is illustrated by first computing a predicted tracking trajectory based on the estimated Q_{est} and then comparing the predicted trajectory with the trajectories collected in experiments.

4.1 Experimental set-up

We apply the IOC algorithm developed to identify a model for the locomotion of the human elbow. Our interest in this set-up stem from its potential in the area of personalized rehabilitation robotics. More precisely, for rehabilitation of hemiparesis,[1] studies have shown that high-intensity, repeated training of specific tasks is an effective method to help patients recover some of the lost motor skills [27–31]. In this setting, tracking a reference signal is one of the fundamental rehabilitation tasks, but due to personal differences between the patients the tracking locomotion patterns could vary. This motivates personalized controller design for rehabilitation robots that help people in their rehabilitation therapy. Under the assumption of optimal behavior (cf. [3]), the idea is to use IOC to identify the objective function used by the "healthy" side of the patient when performing certain tracking task. This objective function would reflect personal habits and preferences. The estimated objective function could then be used to design a control strategy for the rehabilitation robot that helps minimize the performance difference between the "healthy" and the "unhealthy" side.

The experimental device used is illustrated in Fig. 1. With this device set-up, the ideal continuous-time dynamics of the

mass is given by

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\hat{A}} x + \underbrace{\begin{bmatrix} 0 \\ \frac{1}{I} \end{bmatrix}}_{\hat{B}} u,$$

where $x = [x_1, x_2]^T$, and where x_1, x_2, $u \in \mathbb{R}$ denote the angular position, the angular velocity, and the torque applied to the axis, respectively. Moreover, $I = m\ell^2$ denotes the moment of inertia, where $m = 0.2$ and $\ell = 0.255$ denote the mass (in kilograms) and the distance (in meters) between the center of the mass and the axis, respectively. Since measurements are taken every 0.05 seconds, we discretize the ideal system dynamics in time and obtain the discrete time system (A, B), where $A = e^{\hat{A}\Delta t}$, $B = \int_0^{\Delta t} e^{\hat{A}t}dt\hat{B}$, and where $\Delta t = 0.05$. However, due to the existence of friction in the bearing (albeit small), as well as a certain amount of intrinsic randomness in human locomotion, in the discrete-time model we add a process noise $\boldsymbol{w}_t$ to the control input and hence get a discrete-time system in the form of (1b).

Notably, the aforementioned discrete-time dynamics is mathematically the same as the model for pushing a point mass along a one-dimensional line. Hence, in the experimental set-up the state $\boldsymbol{x}_t$ is illustrated on a computer screen by a corresponding block sliding back and forth in one dimension. The movement of the block can be controlled by operating the experiment device, as illustrated in Fig. 1. In the elbow rehabilitation tracking task, we set $\nu_2 = 120$ and generate an oscillating reference signal $x^r_{1:120}$ as

$$x^r_{t+1} = Ax^r_t + Bu^r_t, \quad t = 1 : \nu_2 = 120, \qquad (15)$$
$$x^r_1 = [0, -0.5]^T, \quad u^r_t = 0.01 \sin(\frac{\pi}{40} \cdot t).$$

This defines the behavior of the "target" block to track on the computer screen. The reference signal is shown multiple times to the human before the test starts, so that he/she is totally aware of the reference signal. Once the test starts, the reference signal is played on the computer screen. The human can choose when to start the tracking, and the data is recorded until the final time instant $\nu_2 = 120$.

Furthermore, to measure the mean and covariance of the process noise, i.e., to estimate $\mathbb{E}[\boldsymbol{w}_t]$ and Σ_w, we construct another reference signal. This signal moves at a constant speed, and we let the human track it. The rationale of such set-up lies in the fact that, once the task of tracking a constant velocity reference signal gets close to a "steady state", the control input $\boldsymbol{u}_t$ should be close to zero. Hence any fluctuations of the velocity realizations $x_{t,2}$ can be modeled as caused merely by the process noise $\boldsymbol{w}_t$. Thus, the realizations of the process noise $\boldsymbol{w}_t$ are calculated by $w^i_t = B^\dagger(x^i_{t+1} - Ax^i_t)$, where $\dagger$ denotes the Moore-Penrose generalized inverse. We use the standard empirical estimation to estimate the mean $\mathbb{E}[\boldsymbol{w}_t]$ as well as the covariance Σ_w, and obtain $\mathbb{E}[\boldsymbol{w}_t] \approx 3.3667 \times 10^{-4} \approx 0$ and $\Sigma_w \approx 6.8062 \times 10^{-4}$.

Finally, to solve the estimation problem (14) we implement the problem in Matlab using YALMIP [32] and then solve it numerically using MOSEK [33].

---

[1] Hemiparesis is a medical condition of reduced mobility or weakness in one side of the body that can occur, e.g., after stroke or trauma [26].

4.2 Simulation results

To show the performance of our IOC algorithm and also illustrate the statistical consistency, we use the same parameters as in the experiment set-up described in Sec. 4.1 and simulate $M = 5000$ observed trajectories. The length of each trajectory, denoted N^i, is generated by sampling from a uniform distribution on the integers in $[80, 120]$. For each simulated trajectory, the angular position from which the tracking starts is generated by $x_{120-N^i+1,1} = x^r_{120-N^i+1,1} + \varepsilon^i$, where ε^i is sampled from a uniform distribution on $[-\frac{\pi}{6}, \frac{\pi}{6}]$, and the angular velocity from which the tracking starts is set to zero, i.e., $x_{120-N^i+1,2} = 0$. The "true" parameter $\bar{Q}$ in the objective function is set to $0.01I$. In the reconstruction, we set φ in Assumption 5 to 50. Relative errors of the estimates obtained using the developed algorithm are shown in Fig. 2, and as can be seen the relative estimation error of the parameter decreases as M increases, illustrating the statistical consistency.
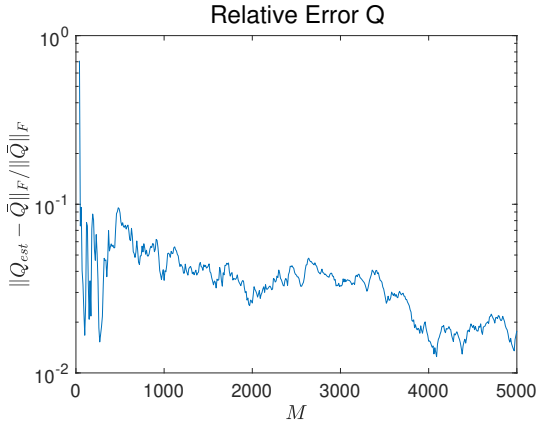

Fig. 2: The relative estimation error vs. the number of observed trajectories M for the simulation experiment described in Sec. 4.2.

4.3 Experimental results

We apply the developed algorithm to data from experiments using the experiment set-up described in Sec. 4.1. More precisely, we collect $M = 500$ trajectories from such tracking experiments. The parameter estimate Q_{est} obtained, based on the human tracking data, is

$$Q_{est} = \begin{bmatrix} 0.0043 & 0.0017 \\ 0.0017 & 0.0007 \end{bmatrix}.$$

To verify the validity of this estimate, we compare the predicted tracking behavior with measured tracking behavior on two new reference signals. In particular, the new reference signals are generated as in (15), but where the first one is generated by $u^{r_1}_t = 0.01\cos(\frac{\pi}{40}t)$ and $x^{r_1}_1 = [0, 0.05]^T$, and second is generated by $u^{r_2}_t = 0$ and $x^{r_2}_1 = [0, 0.2]^T$. Notably, the new reference signals are *different* from the one that is used to collect the training data used in the IOC algorithm. With these new reference signals, tracking experiments are performed and data is collected. Nevertheless, in order to evaluate the prediction accuracy despite the process noise, in each tracking experiment performed we let the human start the tracking at the same angular position each

time (and with a zero angular velocity). To obtain the predicted tracking behavior, we numerically solve the forward optimal control problem using the estimated parameter Q_{est} and the two new reference signals, but without the process noise. The process noise is removed from the generation of the prediction in order to facilitate the comparison, since the form of the control signal in (3) is not explicitly dependent on the process noise. For each reference signal and planing horizon we hence obtain a predicted trajectory $x^{pred}_{\nu_2-N+1:\nu_2}$. The predicted trajectory and the actual collected trajectories, for both reference signals, are illustrated in Fig. 3. As can be seen, the prediction based on estimated parameter Q_{est} is quite accurate in the sense that the actual human tracking locomotion is well-predicted by the trajectories computed by the corresponding Q_{est}. This indicates that the model proposed in the paper describes actual human tracking locomotion well, and thus provide a good model for rehabilitation robots controller design in personalized rehabilitation.
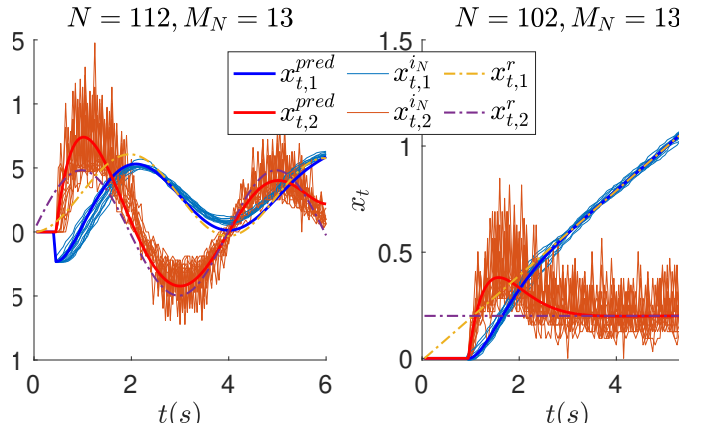

Fig. 3: Illustration of the predicted signals and the measure signals, as described in Sec. 4.3. Due to space limitation, we only show the cases of the planing horizon $N = 112$ (for which $M_N = 13$) for $x^{r_1}_t$ and the planing horizon $N = 102$ (for which $M_N = 13$) for $x^{r_2}_t$. Similar results are obtained for the other planning horizon lengths.

5 Conclusion and future work

IOC provides a powerful theoretical framework for modelling and identifying expert systems. Namely, it enables data-driven objective function design in optimal control, where "expert data" can be used to develop a control strategy that suits the contextual environment. As a conclusion:

1) We develop a statistically consistent IOC algorithm for linear-quadratic tracking problems with random planing horizons, which has not been considered before. Treating the planning horizon as a random variable allows for a systematic way to handle observed data records of different time lengths, while still retaining guarantees about statistical consistency. The statistical consistency ensure the robustness of the estimate and has not been considered in existing more general IOC frameworks, such as [9, 13–17].

2) The proposed IOC algorithm is based on convex optimization. The convexity guarantees that the estimate obtained is the global optima, and hence that the theoretical statistical consistency is actually achieved in real

practice. This has not been achieved in existing work, such as [10, 16, 17].

3) The proposed IOC algorithm is implemented and verified on both simulated and real data. The former data is used to demonstrate the statistical consistency in practice, while the latter is used to demonstrate the capabilities of the method in applications. More precisely, the data is collected in an experimental set-up with a human tracking task motivated from rehabilitation robotics, and the results show that the actual human tracking locomotion is well-predicted by the corresponding estimated objective function – even when the reference signals are different from the one that is used for training.

Future work includes extending the results to more sophisticated nonlinear IOC problems and applying the results to the controller design for actual rehabilitation robots.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal control: linear quadratic methods*. Mineola, NY: Dover publications, 2007.

[2] D. P. Bertsekas, *Dynamic programming and optimal control*, vol. 1. Belmont, MA: Athena scientific, 2nd ed., 2000.

[3] R. M. Alexander, *Optima for animals*. Princeton, NJ: Princeton University Press, 1996.

[4] R. E. Kalman, "When is a linear control system optimal?," *Journal of Basic Engineering*, vol. 86, no. 1, pp. 51–60, 1964.

[5] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. Philadelphia, PA: SIAM, 1994.

[6] M. C. Priess, R. Conway, J. Choi, J. M. Popovich, and C. Radcliffe, "Solutions to the inverse LQR problem with application to biological systems analysis," *IEEE Transactions on control systems technology*, vol. 23, no. 2, pp. 770–777, 2014.

[7] Y. Li, H. Zhang, Y. Yao, and X. Hu, "A convex optimization approach to inverse optimal control," in *2018 37th Chinese Control Conference (CCC)*, pp. 257–262, IEEE, 2018.

[8] Y. Li, Y. Yao, and X. Hu, "Continuous-time inverse quadratic optimal control problem," *Automatica*, vol. 117, p. 108977, 2020.

[9] A. Keshavarz, Y. Wang, and S. Boyd, "Imputing a convex objective function," in *2011 IEEE international symposium on intelligent control*, pp. 613–619, IEEE, 2011.

[10] H. Zhang, J. Umenberger, and X. Hu, "Inverse optimal control for discrete-time finite-horizon linear quadratic regulators," *Automatica*, vol. 110, p. 108593, 2019.

[11] C. Yu, Y. Li, H. Fang, and J. Chen, "System identification approach for inverse optimal control of finite-horizon linear quadratic regulators," *Automatica*, vol. 129, p. 109636, 2021.

[12] H. Zhang and A. Ringh, "Inverse linear-quadratic discrete-time finite-horizon optimal control for indistinguishable agents: A convex optimization approach," *arXiv preprint arXiv:2109.01040v1*, 2021.

[13] K. Hatz, J. P. Schloder, and H. G. Bock, "Estimating parameters in optimal control problems," *SIAM Journal on Scientific Computing*, vol. 34, no. 3, pp. A1707–A1728, 2012.

[14] E. Pauwels, D. Henrion, and J.-B. Lasserre, "Linear conic optimization for inverse optimal control," *SIAM Journal on Control and Optimization*, vol. 54, no. 3, pp. 1798–1825, 2016.

[15] J. Rouot and J.-B. Lasserre, "On inverse optimal control via polynomial optimization," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 721–726, IEEE, 2017.

[16] T. L. Molloy, J. J. Ford, and T. Perez, "Finite-horizon inverse optimal control for discrete-time nonlinear systems," *Automatica*, vol. 87, pp. 442–446, 2018.

[17] T. L. Molloy, J. J. Ford, and T. Perez, "Online inverse optimal control for control-constrained discrete-time systems on finite and infinite horizons," *Automatica*, vol. 120, p. 109109, 2020.

[18] C. Finn, S. Levine, and P. Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization," in *International conference on machine learning*, pp. 49–58, PMLR, 2016.

[19] F. Köpf, J. Inga, S. Rothfuß, M. Flad, and S. Hohmann, "Inverse reinforcement learning for identification in linear-quadratic dynamic games," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14902–14908, 2017.

[20] K. Mombaur, A. Truong, and J.-P. Laumond, "From human to humanoid locomotion – an inverse optimal control approach," *Autonomous robots*, vol. 28, no. 3, pp. 369–383, 2010.

[21] K. Westermann, J. F.-S. Lin, and D. Kulić, "Inverse optimal control with time-varying objectives: application to human jumping movement analysis," *Scientific reports*, vol. 10, no. 1, pp. 1–15, 2020.

[22] O. Kallenberg, *Foundations of modern probability*. Springer, 1997.

[23] R. A. Horn and C. R. Johnson, *Matrix analysis*. New York, NY: Cambridge university press, 2013.

[24] R. I. Jennrich, "Asymptotic properties of non-linear least squares estimators," *The Annals of Mathematical Statistics*, vol. 40, no. 2, pp. 633–643, 1969.

[25] A. W. van der Vaart, *Asymptotic statistics*. Cambridge, United Kingdom: Cambridge university press, 1998.

[26] American Heart Association, "Hemiparesis." [Online] https://www.stroke.org/en/about-stroke/effects-of-stroke/physical-effects-of-stroke/physical-impact/hemiparesis. Accessed: 2022-02-21.

[27] C. Bütefisch, H. Hummelsheim, P. Denzler, and K.-H. Mauritz, "Repetitive training of isolated movements improves the outcome of motor rehabilitation of the centrally paretic hand," *Journal of the neurological sciences*, vol. 130, no. 1, pp. 59–68, 1995.

[28] G. Kwakkel, R. C. Wagenaar, T. W. Koelman, G. J. Lankhorst, and J. C. Koetsier, "Effects of intensity of rehabilitation after stroke: a research synthesis," *Stroke*, vol. 28, no. 8, pp. 1550–1556, 1997.

[29] G. Kwakkel, B. J. K. Kollen, and R. C. Wagenaar, "Long term effects of intensity of upper and lower limb training after stroke: a randomised trial," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 72, no. 4, pp. 473–479, 2002.

[30] N. A. Bayona, J. Bitensky, K. Salter, and R. Teasell, "The role of task-specific training in rehabilitation therapies," *Topics in stroke rehabilitation*, vol. 12, no. 3, pp. 58–65, 2005.

[31] N. Hogan, H. I. Krebs, B. Rohrer, and J. J. Palazzolo, "Motions or muscles? Some behavioral factors underlying robotic," *Journal of Rehabilitation Research & Development*, vol. 43, no. 5-7, p. 601, 2006.

[32] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *2004 IEEE international conference on robotics and automation (IEEE Cat. No. 04CH37508)*, pp. 284–289, IEEE, 2004.

[33] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019.