Online Stochastic Optimization for Unknown Linear Systems: Data-Driven Controller Synthesis and Analysis

Gianluca Bianchin, Miguel Vaquero, Jorge Cortés, and Emiliano Dall'Anese

***Abstract*—This paper proposes a data-driven control framework to regulate an unknown stochastic linear dynamical system to the solution of a stochastic convex optimization problem. Despite the centrality of this problem, most of the available methods critically rely on a precise knowledge of the system dynamics, thus requiring offline system identification. To solve the control problem, we first show that the steady-state gain of the transfer function of a linear system can be computed directly from historical data generated by the open-loop system, thus overcoming the need to first identify the full system dynamics. We leverage this data-driven representation of the steady-state gain to design a controller, which is inspired by stochastic gradient descent methods, to regulate the system to the solution of the prescribed optimization problem. A distinguishing feature of our method is that it does not require any knowledge of the system dynamics or of the possibly time-varying disturbances affecting them (or their distributions). Our technical analysis combines concepts from behavioral system theory, stochastic optimization with decision-dependent distributions, and Lyapunov stability. We illustrate the applicability of the framework in a case study for mobility-on-demand ride service scheduling in Manhattan.**

I. Introduction

This paper focuses on the design of output-feedback controllers to regulate a discrete-time linear time-invariant system to an equilibrium point that is the solution to a convex optimization problem. Due to the presence of unknown and unmeasurable disturbance terms entering the system, the solution of the optimization problem cannot be computed explicitly; thus, the control inputs must be adapted online based on the instantaneous output feedback measured from the system. Our controller synthesis is inspired by principled optimization methods, suitably modified to account for unmeasurable disturbance terms and for the presence of dynamics in the system to control. These problems are central in application domains such as power grids [2], [3], transportation systems [4], robotics [5], and control of epidemics [6], where the target optimization problem encodes desired performance objectives and constraints (possibly dynamic and time-varying) of the system at equilibrium. Within this broad context, the focus of this work is on cases where the target equilibrium point is defined according to a *stochastic* optimization problem and on proposing a new approach to controller synthesis that is *based on data*, thus enabling controller synthesis in cases where the underlying dynamical system is *stochastic* and *unknown*.

Most of the recent literature on online optimization for dynamical systems critically relies on the availability of a model describing the dynamics of the system to control. For instance, the authors in [7] model the system to control as an algebraic map and assume that such a map is known; the backstepping technique proposed in [8] requires exact knowledge of the system dynamics to recast control laws derived in terms of the system output into explicit control laws for the system inputs; the control techniques in [3], [4], [9] focus on systems with linear dynamics and require knowledge of the system's steady-state gain of the transfer function; the methods [10], [11] account for systems with nonlinear dynamics and require knowledge of the (Jacobian of) the system's steady-state map. Unfortunately, perfect system knowledge is rarely available in practice – especially when exogenous disturbances are not observable and/or inputs are not persistently exciting – because maintaining and refining full system models often require ad-hoc system-identification phases. In lieu of system-based controller synthesis, data-driven controllers can be fully synthesized by leveraging data from past trajectories. To the best of our knowledge, the design of optimization-based controllers that overcome the need to identify the system dynamics is still lacking in the literature.

Related Work. Data-driven control methods exploit the ability to express future trajectories of a linear system in terms of a sufficiently-rich past trajectory, as shown by the fundamental lemma [12]. This result, developed in the context of behavioral theory, has enabled the synthesis of several types of controllers, including static feedback controllers [13]–[15], model predictive controllers [16], [17], minimum-energy control laws [18], to solve trajectory tracking problems [19], distributed control problems [20], and recent extensions account for systems with nonlinear dynamics [21], [22].

The line of research on online convex optimization [23] is also related to this work. Several works recently applied online convex optimization tools to control plants modeled as algebraic maps [24]–[26] (describing systems whose dynamics are infinitely-fast). When the dynamics are non-negligible, LTI systems are considered in [2]–[4], [27], stable nonlinear systems in [11], [28], switching systems in [9], and distributed multi-agent systems in [8], [29]. All these works consider continuous-time deterministic dynamics and deterministic optimization problems. In contrast, in this work, we focus on discrete-time stochastic dynamical systems and stochastic optimization problems. Discrete dynamics were the focus of [5], which however does not account for the presence of disturbances affecting the system. To the best of our knowledge,

A preliminary version of this paper [1] has appeared at the 2021 IEEE Conf. on Decision and Control. This work was supported by NSF through Awards CMMI 2044946 and 2044900, and by the National Renewable Energy Laboratory through the subcontract UGA-0-41026-148.

G. Bianchin is with the ICTEAM and the Department of Mathematical Engineering, University of Louvain, Belgium. E. Dall'Anese is with the Department of Electrical, Computer, and Energy Engineering, University of Colorado Boulder, USA. M. Vaquero is with the School of Human Sciences and Technology, IE University. J. Cortés is with the Department of Mechanical and Aerospace Engineering, University of California San Diego, USA.

data-driven implementations of online optimization controllers have not been explored yet. A notable exception is [30], which however does not account for the presence of noise, and results are limited to regret analysis. In this work, we show that when the system is affected by an unknown disturbance, the distributions of the random variables that characterize the optimization are parametrized by the decision variables, thus leading to a stochastic optimization problem *with decision-dependent distributions*, as studied in [31]–[33]. In this work, we build upon this class of problems, but accounting for two additional complexities: the online nature of the optimization method and the coupling with a dynamical system.

Contributions. The contribution of this work is fourfold[1]. First, we show that the steady-state gain of the transfer function of a linear system can be computed from non-steady-state, open-loop, and finite-length input-output trajectories generated by the system (without any knowledge or estimation of the system parameters). When the historical data is affected by unknown noise terms, our techniques return a class of steady-state transfer functions that are compatible with the noisy data and, in this case, we explicitly characterize the approximation error. We remark that, differently from [19], [34]–[36], our framework accounts for disturbances that enter both the output equation and the model equation. Second, we leverage this data-driven representation of the system to propose a control method (inspired by an online version of the stochastic projected gradient-descent algorithm [37]) that regulates such system to an equilibrium point that is the solution to a stochastic optimization problem. Our framework shows that when the historical data is affected by noise and the transfer function is known only approximately, the control problem becomes a stochastic optimization problem *with decision-dependent distribution* [33], whereby changes in the control variable induce a shift in the distribution of the underlying random variables. This is a class of problems whose direct solution is in general difficult to compute [31] and thus, to bypass this hurdle, we introduce a new notion of optimal point, named *stable optimizer* [31], [32], Third, we analyze the convergence and performance properties of the proposed controller, showing that is contractive (in expectation) with respect to the stable optimizer and explicitly quantifying its transient performance. More generally, our results show that the controller regulates the system towards a stable optimizer up to an asymptotic error that depends on the time variability of the optimization problem. Fourth, we apply the control method to study a real-time fleet management problem, where a ride service provider seeks to maximize its profit by dispatching its fleet while serving ride requests from its customers. We demonstrate the applicability and benefits of our methods numerically on a real network and demand data.

Organization. Section II presents some basic notions used in the work; Section III formalizes the problem of interest; in Section IV, we discuss our data-driven technique to compute the the transfer function of linear systems. Section V illustrates our controller synthesis method; in Section VI, we introduce the notion of a stable optimizer to tackle optimization problems with decision-dependent distributions, and Section VII presents the controller analysis. Section VIII illustrates an application of the method to ride service scheduling and Section IX concludes the paper.

[1]This paper generalizes the preliminary work [1] in several directions: we consider stochastic optimization problems (rather than deterministic), we account for noisy training data, and we only require systems to be observable (instead of relying on direct state measurements), we provide an explicit contraction bound, and we illustrate applicability on a ride-service scheduling problem.

II. Preliminaries

In this section, we outline the notation and introduce some preliminary concepts used throughout the paper.

Notation. Given a symmetric matrix $M \in \mathbb{R}^{n\times n}$, $\underline{\lambda}(M)$ and $\bar{\lambda}(M)$ denote its smallest and largest eigenvalue, respectively; $M \succ 0$ indicates that M is positive definite. Given a matrix $M \in \mathbb{R}^{n\times m}$, $M^\dagger$ denotes its Moore-Penrose inverse. For a vector $u \in \mathbb{R}^n$, we denote by $\|u\|$ the Euclidean norm of u and by $u^\top$ its transpose. For vectors $u \in \mathbb{R}^n, w \in \mathbb{R}^m$, we use the shorthand notation $(u, w) \in \mathbb{R}^{n+m}$ for their vector concatenation.

Persistency of Excitation. We next recall some useful facts on behavioral system theory from [12]. For a signal $k \mapsto z_k \in \mathbb{R}^\sigma$, $k \in \mathbb{Z}$, we denote the vectorization of z restricted to the interval $[k, k+T]$, $T \in \mathbb{Z}_{>0}$, by $z_{[k,k+T]} := (z_k, \dots, z_{k+T})$. Given $z_{[0,T-1]}$, $t \le T$, and $q \le T - t + 1$, we let $Z_{t,q}$ denote the Hankel matrix of length t associated with $z_{[0,T-1]}$:

$$Z_{t,q} = \begin{bmatrix} z_0 & z_1 & \dots & z_{q-1} \\ z_1 & z_2 & \dots & z_q \\ \vdots & \vdots & \ddots & \vdots \\ z_{t-1} & z_t & \dots & z_{q+t-2} \end{bmatrix} \in \mathbb{R}^{\sigma t\times q}.$$

Moreover, we use $[Z_{t,q}]_i$, $i \in \{1, \dots, t\}$ to denote the i-th block-row of $Z_{t,q}$, namely, $[Z_{t,q}]_i = [z_{i-1},\ z_i,\ \dots,\ z_{i+q-2}]$.

Definition 2.1: ***(Persistently Exciting Signal [12])*** The signal $z_{[0,T-1]}$, $z_k \in \mathbb{R}^\sigma$ for all $k \in \{0, \dots, T-1\}$, is persistently exciting of order t if $Z_{t,q}$ has full row rank σt.□

We note that persistence of excitation implicitly requires $q \ge \sigma t$ (which in turns requires $T \ge (\sigma + 1)t - 1$).

Consider the linear dynamical system

$$x_{k+1} = Ax_k + Bu_k, \qquad y_k = Cx_k + Du_k, \tag{1}$$

with $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$. Let $\mathcal{C}_\theta := [B, AB, A^2B, \dots, A^{\theta-1}B]$ and $\mathcal{O}_\nu := [C^\mathsf{T}, A^\mathsf{T}C^\mathsf{T}, \dots, (A^\mathsf{T})^{\nu-1}C^\mathsf{T}]^\mathsf{T}$ denote the controllability and observability matrices of (1), respectively. The system is controllable if $\mathrm{rank}(\mathcal{C}_\theta) = n$ for some $\theta \in \mathbb{Z}_{>0}$, and it is observable if $\mathrm{rank}(\mathcal{O}_\nu) = n$ for some $\nu \in \mathbb{Z}_{>0}$. The smallest integers θ, ν, that satisfy the above conditions are the controllability and observability indices, respectively.

We next recall the following properties of (1) when its inputs are persistently exciting.

Lemma 2.2: ***(Fundamental Lemma [12, Corollary 2])*** Assume (1) is controllable, let $(u_{[0,T-1]}, x_{[0,T-1]})$, $T \in \mathbb{Z}_{>0}$,

be an input-state trajectory of (1). If $u_{[0,T-1]}$ is persistently exciting of order $n+L$, then:

$$\operatorname{rank}\begin{bmatrix} U_{L,q} \\ X_{1,q} \end{bmatrix} = Lm + n,$$

where $U_{L,q}$ and $X_{1,q}$ denote the Hankel matrices associated with $u_{[0,T-1]}$ and $x_{[0,T-1]}$, respectively. □

Lemma 2.3: ***(Data Characterizes Full Behavior [12, Thm. 1])*** Assume (1) is controllable and observable, let $(u_{[0,T-1]}, y_{[0,T-1]})$, $T \in \mathbb{Z}_{>0}$, be an input-output trajectory of (2). If $u_{[0,T-1]}$ is persistently exciting of order $n+L$, then any pair of L-long signals $(\tilde{u}_{[0,L-1]}, \tilde{y}_{[0,L-1]})$ is an input-output trajectory of (2) if and only if there exists $\alpha \in \mathbb{R}^q$ such that

$$\begin{bmatrix} \tilde{u}_{[0,L-1]} \\ \tilde{y}_{[0,L-1]} \end{bmatrix} = \begin{bmatrix} U_{L,q} \\ Y_{L,q} \end{bmatrix} \alpha,$$

where $U_{L,q}$ and $Y_{L,q}$ denote the Hankel matrices associated with $u_{[0,T-1]}$ and $y_{[0,T-1]}$, respectively. □

In words, persistently exciting signals generate output trajectories that can be used to express any other trajectory.

Probability Theory. Let $(\Omega, \mathcal{F}, P)$ be a probability space and z be a random variable mapping this space to $(\mathbb{R}^d, B_\sigma(\mathbb{R}^m))$, where $B_\sigma(\mathbb{R}^d)$ is the Borel σ-algebra on $\mathbb{R}^d$. Let $\mathcal{P}$ be the distribution of z and $\Xi \subseteq \mathbb{R}^m$ be the support of $\mathcal{P}$. We use $z \sim \mathcal{P}$ to denote that z is distributed according to $\mathcal{P}$, and $\mathbb{E}_{z\sim\mathcal{P}}[\cdot]$ to denote the expectation under $\mathcal{P}$. Let $\mathcal{M}(\Xi)$ be the space of all probability distributions supported on Ξ with finite first moment, i.e., $\mathbb{E}_{z\sim\mathcal{P}}[\|z\|] = \int_\Xi \|z\|\mathcal{P}(dz) < \infty$ for all $\mathcal{P} \in \mathcal{M}(\Xi)$. The Wasserstein-1 metric is:

$$W_1(\mathcal{P}_1, \mathcal{P}_2) := \inf_{\Pi \in \mathcal{H}(\mathcal{P}_1, \mathcal{P}_2)} \left\{ \int_{\Xi^2} \|z_1 - z_2\| \Pi(dz_1, dz_2) \right\},$$

where $\mathcal{H}(\mathcal{P}_1, \mathcal{P}_2)$ is the set of all joint distributions with marginals $\mathcal{P}_1$ and $\mathcal{P}_2$. By interpreting the decision function $\Pi(\mathcal{P}_1, \mathcal{P}_2)$ as a transportation plan for moving a mass distribution described by $\mathcal{P}_1$ to another one described by $\mathcal{P}_2$, the Wasserstein distance $W_1(\mathcal{P}_1, \mathcal{P}_2)$ represents the cost of an optimal mass transportation plan, where the transportation costs are described by the 1-Euclidean norm.

Theorem 2.4: ***(Kantorovich-Rubinstein, [38])*** For any pair of distributions $\mathcal{P}_1, \mathcal{P}_2 \in \mathcal{M}(\Xi)$, the following holds

$$W_1(\mathcal{P}_1, \mathcal{P}_2) = \sup_{g \in \mathcal{L}_1} \left\{ \mathbb{E}_{z_1\sim\mathcal{P}_1}[g(z_1)] - \mathbb{E}_{z_2\sim\mathcal{P}_2}[g(z_2)] \right\},$$

where $\mathcal{L}_1$ denotes the space of all 1-Lipschitz functions, i.e., $\mathcal{L}_1 := \{g : \Xi \to \mathbb{R}^m \ : \ \|g(z_1) - g(z_2)\| \le \|z_1 - z_2\|\}$. □

The following result is instrumental for our analysis. We provide a short proof for completeness.

Lemma 2.5: ***(Deviation Between Expectations [31, Lemma C.4])*** Let $f : \mathbb{R}^n \to \mathbb{R}^d$ be L-Lipschitz continuous. Then, for any pair of distributions $\mathcal{P}_1, \mathcal{P}_2 \in \mathcal{M}(\Xi)$,

$$\|\mathbb{E}_{z_1\sim\mathcal{P}_1}[f(z_1)] - \mathbb{E}_{z_2\sim\mathcal{P}_2}[f(z_2)]\| \le L W_1(\mathcal{P}_1, \mathcal{P}_2).$$

Proof: Let $v \in \mathbb{R}^d$ be any unit vector and let $g(z) := v^\mathsf{T} f(z)$. By assumption, $g(z)$ is L-Lipschitz continuous and therefore:

$$\begin{aligned} v^\mathsf{T}(&\mathbb{E}_{z_1\sim\mathcal{P}_1}[f(z_1)] - \mathbb{E}_{z_2\sim\mathcal{P}_2}[f(z_2)]) \\ &= \mathbb{E}_{z_1\sim\mathcal{P}_1}\left[v^\mathsf{T} f(z_1)\right] - \mathbb{E}_{z_2\sim\mathcal{P}_2}\left[v^\mathsf{T} f(z_2)\right]) \\ &= \mathbb{E}_{z\sim\mathcal{Z}(\alpha)}[g(z)] - \mathbb{E}_{z\sim\mathcal{Z}(\beta)}[g(z)] \le L W_1(\mathcal{P}_1, \mathcal{P}_2), \end{aligned}$$

where the last inequality follows by the application of Theorem 2.4. The result then follows by choosing

$$v = \frac{\mathbb{E}_{z_1\sim\mathcal{P}_1}[f(z_1)] - \mathbb{E}_{z_2\sim\mathcal{P}_2}[f(z_2)]}{\|\mathbb{E}_{z_1\sim\mathcal{P}_1}[f(z_1)] - \mathbb{E}_{z_2\sim\mathcal{P}_2}[f(z_2)]\|}.$$ ■

III. Problem Formulation

In this section, we present the problem focus of this work and we discuss a reformulation used for controller synthesis.

A. Steady-State Regulation Problem for Linear Systems

We consider stochastic discrete-time linear system:

$$x_{k+1} = Ax_k + Bu_k + Ew_k, \qquad y_k = Cx_k + Dw_k, \tag{2}$$

where $k \in \mathbb{Z}_{>0}$ is the time index, $x_k \in \mathbb{R}^n$ is the state, $u_k \in \mathbb{R}^m$ is the control decision, $y_k \in \mathbb{R}^p$ is the measurable output, and $\mathbb{R}^r \ni w_k \sim \mathcal{W}_k$ is an exogenous disturbance whose distribution $\mathcal{W}_k$ is unknown for all $k \in \mathbb{Z}_{>0}$.

Assumption 1: ***(System Properties)*** System (2) is controllable and observable. Moreover, A is Schur stable, i.e., for any $Q \succ 0$, there exists $P \succ 0$ such that $A^\mathsf{T} P A - P = -Q$. □

Assumption 1 guarantees that for any pair of (fixed) vectors $\bar{u} \in \mathbb{R}^m$ and $\bar{w} \in \mathbb{R}^r$, (2) admits a unique stable equilibrium point described by the state $(I-A)^{-1}B\bar{u} + (I-A)^{-1}E\bar{w}$. At the equilibrium, the system output is given by:

$$\underbrace{C(I-A)^{-1}B}_{:=G}\,\bar{u} + \underbrace{(D + C(I-A)^{-1}E)}_{:=H}\,\bar{w}, \tag{3}$$

where G denotes the steady-state gain of the transfer function (or steady-state transfer function for short) of the system and H denotes the disturbance-to-output steady-state gain.

In this work, we focus on the problem of controlling, at every time k, the system (2) to the solution of the following time-dependent optimization problem:

$$\min_{\bar{u}} \ \ \mathbb{E}_{w_k\sim\mathcal{W}_k}\left[\phi(\bar{u}, G\bar{u} + Hw_k)\right], \tag{4}$$

where $\phi : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is a cost function that describes losses associated with the steady-state control inputs $\bar{u}$ and steady-state system outputs (given by $G\bar{u} + Hw_k$). Notice that, in (4), w_k parametrizes the optimization problem, which is thus time-dependent. Problem (4) formalizes an optimal equilibrium-selection problem, where the objective is to select, at every time k, an optimal equilibrium point (as specified by the cost $\phi(\cdot,\cdot)$) despite the instantaneous (unknown) noise $w_k \sim \mathcal{W}_k$ affecting the dynamics.

Two important observations on (4) are in order. First, because the distribution $\mathcal{W}_k$ is time-dependent, the solution of the optimization will also be time-dependent. Second, because $\mathcal{W}_k$ is unknown and samples of w_k are unmeasurable, a solution of the problem (4) cannot be computed explicitly.

Remark 1: ***(Relationship with Output Regulation Problems)*** Problem (4) formalizes a steady-state control problem, similar to the well-established output-regulation problem [39], where the objective is to find a control law such that the output of the plant can asymptotically track a prescribed trajectory and reject undesired disturbances. Differently from the classical output-regulation problem, where the target trajectory can be arbitrary and is often generated by an exosystem (i.e., a known autonomous linear model), in our setting target trajectories are specified as the solution of an optimization problem (whose solution is unknown since $\mathcal{W}_k$ is unknown). □

We impose the following regularity assumptions on $\phi(u, y)$.

Assumption 2: ***(Lipschitz and Convexity of Cost Function)***

(a) For any fixed $u \in \mathbb{R}^m$, the map $y \mapsto \phi(u, y)$ is ℓ-Lipschitz continuous.
(b) For any fixed $y \in \mathbb{R}^p$, the map $u \mapsto \nabla\phi(u, y)$ is ℓ_u^∇-Lipschitz continuous, and for any fixed $u \in \mathbb{R}^m$ the map $y \mapsto \nabla\phi(u, y)$ is ℓ_y^∇-Lipschitz continuous.
(c) For any fixed $y \in \mathbb{R}^p$, the map $u \mapsto \phi(u, y)$ is μ-strongly convex, i.e., there exists $\mu \in \mathbb{R}_{>0}$ such that, for all $u, u' \in \mathbb{R}^m$, $\phi(u, y) \geq \phi(u', y) + \nabla\phi(u', y)^\mathsf{T}((u, y) - (u', y)) + \frac{\mu}{2}\|u - u'\|^2$. □

Strong convexity and Lipschitz-type assumptions impose basic conditions on the growth of the cost function often used for the analysis of first-order optimization methods [40]. Under Assumption 2(c), the function $\bar{u} \mapsto \mathbb{E}_{w_k \sim \mathcal{W}_k}[\phi(\bar{u}, G\bar{u} + Hw_k)]$ is also strongly convex [41], and thus (4) admits a unique optimizer u_k^* for each k. Further, it follows from Assumption 1 that the corresponding equilibrium state $x_k^* = (I-A)^{-1}Bu_k^* + (I - A)^{-1}Ew_k$ is unique since the linear map $(I - A)^{-1}$ has empty kernel. Hence, in what follows, we denote by (u_k^*, x_k^*) the unique optimizer associated to (4). We conclude by formally stating the problem studied in this work.

Problem 1: Design a dynamic output-feedback controller of the form $u_{k+1} = \mathcal{C}(u_k, y_k)$ such that, without any prior knowledge of the matrices (A, B, C, D, E) as well as of the noise distributions $\mathcal{W}_k$, the inputs and outputs of (2) converge asymptotically to the time-varying optimizer of (4). □

B. Problem Reformulation for Unknown Dynamics

When the matrices (A, B, C, D, E) are unknown, the maps G and H in (3) are also unknown, and thus the steady-state output of (2) as well as its effect on the optimization problem (4) are unknown. In practice, however, it is often possible to use some basic knowledge of the system (2) to derive an approximation $\hat{G}$ of the map G. (Notice that estimating H is in general a more challenging task since w_k is unmeasurable). In these cases, problem (4) can be rewritten as:

$$u_k^* = \arg\min_{\bar{u}} \quad \mathbb{E}_{z_k \sim \mathcal{Z}_k(\bar{u})}[\phi(\bar{u}, \hat{G}\bar{u} + z_k)], \tag{5}$$

where $z_k := (G - \hat{G})\bar{u} + Hw_k$ is a random variable that encodes the lack of knowledge of the map G as well as of the term Hw_k, and where $\hat{G}$ is treated as a fixed (non stochastic) matrix. We note that in (5) the distribution of z_k is parametrized by the decision variable $\bar{u}$ and, in order to emphasize such dependency, we used the notation[2] $z_k \sim \mathcal{Z}_k(\bar{u})$. From an optimization perspective, seeking a solution of (5) raises three main challenges:

(C1) The closed-form expression for G depends on the system matrices (A, B, C) (cf. (3)), which are unknown. This raises the question of how to construct an approximate map $\hat{G}$ and how to quantify the approximation error.
(C2) Since $\mathcal{W}_k$ is unknown, the distribution $\mathcal{Z}_k(\bar{u})$ is also unknown. Instead, we only have access to evaluations of z_k via the output y_k of the system (2). This calls for the development of controllers that can adjust u_k based on noisy evaluations of the cost function through access to the instantaneous system output y_k.
(C3) Because the distribution of z_k is parametrized by the decision variable $\bar{u}$, the cost function of (5) is nonlinear in $\bar{u}$, thus making the optimization problem (5) intractable for general costs (even when $\phi(u, y)$ is convex).

The subsequent sections address the above challenges. Precisely, we address (C1) in Section IV; we propose a control method that accounts for (C2) in Section V; and in Section VI we address (C3) by introducing an alternative notion of optimality. We put together all these developments in Section VII and provide the technical analysis of the proposed controller.

IV. Data-Driven Computation of the Transfer Function of Linear Systems

In this section, we tackle challenge (C1). To this end, we show that G can be computed from input, output, and noise data generated by (2). When the noise terms are unknown, we show that our method yields an approximation $\hat{G}$ of G.

A. Direct Computation of the Transfer Function from Data

We begin by assuming the availability of a set of historical data $(u_{[0,T]},\ w_{[0,T]}, y_{[0,T]}), T \in \mathbb{Z}_{>0}$, generated by (2). In order to state our result, we define:

$$\begin{aligned} y^{\text{diff}} &:= (y_1 - y_0,\ y_2 - y_1, \dots,\ y_T - y_{T-1}), \\ w^{\text{diff}} &:= (w_1 - w_0,\ w_2 - w_1, \dots,\ w_T - w_{T-1}), \end{aligned} \tag{6}$$

and we let $Y_{\nu,q}^{\text{diff}}$ and $W_{\nu,q}^{\text{diff}}$, respectively, be the associated Hankel matrices. The following result provides a method to compute G via algebraic operations on the historical data.

Theorem 4.1: ***(Data-Driven Characterization of Steady-State Transfer Function)*** Let Assumption 1 be satisfied and let $\nu \in \mathbb{Z}_{>0}$ denote the observability index of (2). Moreover, assume that the concatenated signal $(u_{[0,T-1]}, w_{[0,T-1]})$ is persistently exciting signals of order $n+\nu$, and let $q := T-\nu+1$. The following holds:

(i) There exists $M \in \mathbb{R}^{q \times m}$ such that:

$$\begin{aligned} Y_{\nu,q}^{\text{diff}} M &= 0, & W_{\nu,q}^{\text{diff}} M &= 0, \\ U_{\nu,q} M &= \mathbb{1}_\nu \otimes I_m, & W_{\nu,q} M &= 0, \end{aligned} \tag{7}$$

where $Y_{\nu,q}^{\text{diff}}$ and $W_{\nu,q}^{\text{diff}}$ are defined in (6).

[2]To ease the notation in what follows we denote $\mathbb{E}_{\bar{z} \sim \mathcal{Z}_k(\bar{u})}[\cdot]$ in compact form as $\mathbb{E}_{\mathcal{Z}_k(\bar{u})}[\cdot]$, since the random variable with respect to which the expectation is taken is made clear in the argument.

(ii) For any $M \in \mathbb{R}^{q\times m}$ that satisfies (7), the steady-state transfer function of (2) equals $G = [Y_{\nu,q}]_i M$, for any $i \in \{1,\dots,\nu\}$. □

Proof: (Proof of (i)). Fix a $j \in \{1,\dots,m\}$, let $\bar{u} = (e_j, e_j, \dots) \in \mathbb{R}^{m\nu}$, where $e_j \in \mathbb{R}^m$ denotes the j-th canonical vector, let $\bar{w} = (\mathbb{0}_r, \mathbb{0}_r, \dots) \in \mathbb{R}^{r\nu}$, and let $\bar{y} = (Ge_j, Ge_j, \dots) \in \mathbb{R}^{p\nu}$. Since $(\bar{u}, \bar{w}, \bar{y})$ is an input-output trajectory of (2), Lemma 2.2 guarantees the existence of $m_j \in \mathbb{R}^q$ such that $U_{\nu,q} m_j = \bar{u}$, $W_{\nu,q} m_j = \bar{w}$, and $Y_{\nu,q} m_j = \bar{y}$. By iterating the above reasoning for all $j \in \{1,\dots,m\}$, and by letting the j-th column of M be m_j, we obtain that $U_{\nu,q} M = \mathbb{1}_\theta \otimes I_m$ and $W_{\nu,q} M = 0$. Moreover, since $\bar{y}$, and $\bar{w}$ are constant at all times we conclude that $Y^{\text{diff}}_{\nu,q} M = 0$ and $W^{\text{diff}}_{\nu,q} M = 0$, which proves existence of M.

(Proof of (ii)). The proof of this claim builds upon the following observation. Let $\bar{U} := I_m$, let $\bar{W} := \mathbb{0}_{r\times m}$, let $\bar{X} := (I_n - A)^{-1} B\bar{U} + (I_n - A)^{-1} E\bar{W}$, and let $\bar{Y} := C\bar{X} + D\bar{W}$. Then, by substitution, $\bar{Y}$ satisfies:

$$\bar{Y} = G\bar{U} + H\bar{W} = G. \tag{8}$$

In words, this implies that when the inputs $\bar{U} = I_m$ and $\bar{W} = \mathbb{0}_{r\times m}$ are applied to (2) and the state satisfies $\bar{X} = (I_n - A)^{-1} B\bar{U} + (I_n - A)^{-1} E\bar{W}$, then the system output satisfies $\bar{Y} = G$, namely, it coincides with the steady-state transfer function G.

Building upon this observation, in what follows we show that (7) and (8) are equivalent, in the sense described by Lemma 2.3. Formally, let M be any matrix that satisfies (7). By application of Lemma 2.3, $U_{\nu,q} M = \mathbb{1}_\nu \otimes I_m$ implies that the input applied to (2) is $\bar{U} = I_m$, and $W_{\nu,q} M = 0$ implies that the exogenous disturbance applied to (2) is $\bar{W} := \mathbb{0}_{r\times m}$. Next, we show that the matrix $\bar{Y}$ defined as $\bar{Y} = [Y_{\nu,q}]_i M$ for any $i \in \{1,\dots,\nu\}$ coincides with (8), namely, we will show:

$$\bar{Y} = [Y_{\nu,q}]_i M \quad \Rightarrow \quad \bar{Y} = C\bar{X}, \bar{X} = (I_n - A)^{-1} B I_m. \tag{9}$$

To this aim, we let $\bar{y}_{ij} = [Y_{\nu,q}]_i m_j$ denote the j-th column of $\bar{Y}$, and we define $\bar{x}_{ij} := [X_{\nu,q}]_i m_j$. Notice that $\bar{Y} = C\bar{X}$ follows from $[Y_{\nu,q}]_i = C[X_{\nu,q}]_i$. Thus, we next show that $\bar{X} = (I_n - A)^{-1} B I_m$. The proof is organized into two steps. *(Step 1) We show that* $\bar{y}_{i,j} = \bar{y}_{i+1,j}$. By using $Y^{\text{diff}}_{\nu,q} M = 0$:

$$\begin{aligned} 0 &= [Y^{\text{diff}}_{\nu,q}]_i m_j \\ &= C \begin{bmatrix} (A - I_n) & B & E \end{bmatrix} \begin{bmatrix} [X_{\nu,q}]_i \\ [U_{\nu,q}]_i \\ [W_{\nu,q}]_i \end{bmatrix} m_j + D[W^{\text{diff}}_{\nu,q}]_i m_j \\ &= C(A - I_n)\bar{x}_{ij} + CBe_j, \end{aligned} \tag{10}$$

where the last inequality follows from $\bar{x}_{ij} := [X_{\nu,q}]_i m_j$, $U_{\nu,q} M = \mathbb{1}_\nu \otimes I_m$, $W_{\nu,q} M = 0$, and $W^{\text{diff}}_{\nu,q} M = 0$. Hence, we conclude that

$$\bar{y}_{ij} = C\bar{x}_{ij} = CA\bar{x}_{ij} + CBe_j. \tag{11}$$

Moreover, since (10) holds for all $i \in \{1,\dots,\nu\}$, Lemma 2.3 guarantees that $\bar{u} = (e_j, \dots, e_j) \in \mathbb{R}^{m\nu}$, $\bar{w} = (\mathbb{0}_r, \dots, \mathbb{0}_r) \in \mathbb{R}^{r\nu}$, $\bar{x} = (\bar{x}_{1j}, \dots, \bar{x}_{\nu j}) \in \mathbb{R}^{n\nu}$, and $\bar{y} = (\bar{y}_{1j}, \dots, \bar{y}_{\nu j}) \in \mathbb{R}^{p\nu}$ is an input-state-output trajectory of the system (2), and thus it satisfies the dynamics:

$$\bar{x}_{i+1,j} = A\bar{x}_{ij} + Be_j, \qquad \bar{y}_{ij} = C\bar{x}_{ij}. \tag{12}$$

By combining (11) with (12) we conclude that:

$$\bar{y}_{ij} = CA\bar{x}_{ij} + CBe_j = C(A\bar{x}_{ij} + Be_j) = C\bar{x}_{i+1,j} = \bar{y}_{i+1,j}.$$

(Step 2) We show that $\bar{x}_{i,j} = (I - A)^{-1} Be_j$. By combining $\bar{y}_{i,j} = \bar{y}_{i+1,j}$ with the dynamics (12) we obtain:

$$CA^k \begin{bmatrix} A - I_n & B \end{bmatrix} \begin{bmatrix} \bar{x}_{0,j} \\ e_j \end{bmatrix} = 0, \text{ for all } k \in \{1,\dots,\nu-1\}.$$

By recalling that (2) is Observable (see Assumption 1), the above identity implies $(A - I)\bar{x}_{0,j} + Be_j = 0$ or, equivalently, $\bar{x}_{0,j} = (I - A)^{-1} Be_j$. By recalling that $\bar{x} = (\bar{x}_{1j}, \dots, \bar{x}_{\nu j}) \in \mathbb{R}^{n\nu}$ represents the state associated with the constant input sequences $\bar{u} = (e_j, \dots, e_j) \in \mathbb{R}^{m\nu}$ and $\bar{w} = (\mathbb{0}_r, \dots, \mathbb{0}_r) \in \mathbb{R}^{r\nu}$ (see (12)), we obtain $x_{i+1,j} = x_{i,j}$ for all $i \in \{1,\dots,\nu-1\}$, which implies that $\bar{x}_{i,j} = (I - A)^{-1} Be_j$ holds for all i, thus proving *Step 2*. Finally, $\bar{X} = (I_n - A)^{-1} B I_m$ follows by iterating the above reasoning for all $j \in \{1,\dots,m\nu\}$. ■

Theorem 4.1 shows that G can be computed from (non-steady-state) historical data generated by the open-loop system (2), and without knowledge of the matrices (A, B, C). We note that the approach proposed in Theorem 4.1 is aligned with the vast body of literature (see, e.g. [12], [30], [42]), where past trajectories are utilized directly (i.e., without resorting on system identification) to parametrize the behavior of the system. The result also offers some insights into the number of required data samples[3]: it shows that the required length of the historical trajectory needed for the computation of G grows linearly with the observability index ν and with the size of the state space n. Two technical observations are in order. First, any M chosen according to (7), in general, depends on the realization of $w_{[0,T-1]}$ and on the choice of $u_{[0,T-1]}$. Second, for any fixed $u_{[0,T-1]}$ and $w_{[0,T-1]}$, in general, there exists an infinite number of choices of M that satisfy (7). Despite M not being unique, Theorem 4.1 guarantees that $[Y_{\nu,q}]_i M$ is unique and independent of the choice of $u_{[0,T-1]}$ and $w_{[0,T-1]}$.

As a direct consequence of Theorem 4.1, G admits the following data-explicit expression, which follows from (7).

Corollary 4.2: ***(Data-based Expression of*** *G)* Under the same assumptions of Theorem 4.1, the steady-state transfer function G admits the following data-based expression:

$$G = [Y_{\nu,q}]_i \begin{bmatrix} Y^{\text{diff}}_{\nu,q} \\ U_{\nu,q} \\ W_{\nu,q} \\ W^{\text{diff}}_{\nu,q} \end{bmatrix}^\dagger \begin{bmatrix} 0 \\ \mathbb{1}_\nu \otimes I_m \\ 0 \\ 0 \end{bmatrix}. \tag{13}$$

□

We conclude with a remark that shows that when $w_{[0,T-1]}$ is constant, G can be computed without knowing the noise terms.

[3]Notice that Theorem 4.1 requires persistence of excitation of the T-long signals $u_{[0,T-1]}$ and $w_{[0,T-1]}$. In addition, constructing the difference signals y^{diff} and w^{diff} requires the collection of one additional sample of the signals $y_{[0,T]}$ and $w_{[0,T]}$ (i.e., $T + 1$ samples).

Remark 2: ***(Noise-Independent Characterization for Constant Disturbance)*** When $w_{[0,T-1]}$ is constant at all times, an adaptation of Theorem 4.1 can be used to determine G without requiring the knowledge of $w_{[0,T-1]}$, as outlined next. Define:

$$d_k := x_{k+1} - x_k, r_k := y_{k+1} - y_k, v_k := u_{k+1} - u_k, \quad (14)$$

then, (2) can be rewritten in velocity form as follows:

$$d_{k+1} = Ad_k + Bv_k, \qquad r_k = Cd_k. \quad (15)$$

Thus, by letting $R_{\nu,q}$ and $V_{\nu,q}$ denote the Hankel matrices of the signals $r_{0,T-1}$ and $v_{0,T-1}$, respectively, and $R^{\text{diff}}_{\nu,q}$ be the Hankel matrix of $[r_1 - r_0,\ r_2 - r_1,\ \dots\ r_T - r_{T-1}]$, then, Theorem 4.1 guarantees that the steady-state transfer function of (2) equals $G = [R_{\nu,q}]_i M$, for any $i \in \{1, \dots, \nu\}$, where $q := T - \nu + 1$ and $M \in \mathbb{R}^{q \times m\nu}$ satisfies

$$R^{\text{diff}}_{\nu,q} M = 0, \qquad V_{\nu,q} M = \mathbb{1}_\nu \otimes I_m. \quad (16)$$

Notice that, the Hankel matrices $R_{1,q}, V_{1,q}$, and $R^{\text{diff}}_{1,q}$ can be computed directly from an input-output trajectory of (2) by processing the data[4] as described by (14). Finally, we notice that the only required assumption to apply Theorem 4.1 in this case is that $u_{[0,T]}$ is persistently exciting of order $n + \nu$. Indeed, persistency of excitations of $u_{[0,T]}$ implies persistency of excitations of $v_{[0,T-1]}$ of the same order. This is because the columns of $V_{\nu,q}$ are obtained by subtracting disjoint pairs of columns of $U_{\nu,q+1}$, which are linearly independent. □

B. Uncertainty Characterization

While Theorem 4.1 provides a way to compute G from data, it requires knowledge of $w_{[0,T-1]}$, which is impractical when the disturbance is unmeasurable. To this end, we next construct upon Theorem 4.1 to derive a characterization of all the transfer functions $\hat{G}$ that are consistent with the noisy data.

Proposition 4.3: ***(Error Characterization)*** Let Assumption 1 hold and let $\nu \in \mathbb{Z}_{>0}$ denote the observability index of (2). Moreover, assume that the concatenated signal $(u_{[0,T-1]}, w_{[0,T-1]})$ is persistently exciting signals of order $n + \nu$, and let $q := T - \nu + 1$. Assume $\hat{M} \in \mathbb{R}^{q \times m\nu}$ is any matrix that satisfies:

$$Y^{\text{diff}}_{\nu,q} \hat{M} = 0, \qquad U_{\nu,q} \hat{M} = \mathbb{1}_\nu \otimes I_m, \quad (17)$$

where $Y^{\text{diff}}_{\nu,q}$ is defined according to (6). If $\hat{G}$ is computed as $\hat{G} := [Y_{\nu,q}]_i \hat{M}$, for any $i \in \{1, \dots, \nu\}$, then

$$\begin{aligned} \hat{G} - G = {} & CA([X_{\nu,q}]_i \hat{M} - (I - A)^{-1} B) \quad (18) \\ & + CE[W_{\nu,q}]_i \hat{M} + D([W^{\text{diff}}_{\nu,q}]_i + [W_{\nu,q}]_i)\hat{M}. \end{aligned}$$

Proof: Let M be any matrix as in (7) and $\hat{M}$ be any matrix as in (17). By noting that $\hat{G} - G = [Y_{\nu,q}]_i(\hat{M} - M)$, we will prove this claim by showing that $[Y_{\nu,q}]_i(\hat{M} - M)$ equals the right-hand side of (18). By using $[Y^{\text{diff}}_{\nu,q}]_i M = 0$, $[Y^{\text{diff}}_{\nu,q}]_i \hat{M} = 0$, and by recalling that $[Y^{\text{diff}}_{\nu,q}]_i = C(A - I)[X_{\nu,q}]_i + CB[U_{\nu,q}]_i + CE[W_{\nu,q}]_i + D[W^{\text{diff}}_{\nu,q}]_i$:

$$\begin{aligned} 0 = {} & [Y^{\text{diff}}_{\nu,q}]_i (\hat{M} - M) \quad (19) \\ = {} & C(A - I)[X_{\nu,q}]_i (\hat{M} - M) + CB[U_{\nu,q}]_i (\hat{M} - M) \\ & + CE[W_{\nu,q}]_i (\hat{M} - M) + D[W^{\text{diff}}_{\nu,q}]_i (\hat{M} - M), \\ = {} & C(A - I)[X_{\nu,q}]_i (\hat{M} - M) + CE[W_{\nu,q}]_i \hat{M} + D[W^{\text{diff}}_{\nu,q}]_i \hat{M}, \end{aligned}$$

where we used $[U_{\nu,q}]_i \hat{M} = [U_{\nu,q}]_i M$, $[W_{\nu,q}]_i M = 0$, and $[W^{\text{diff}}_{\nu,q}]_i M = 0$. Next, by recalling that $[Y_{\nu,q}]_i(\hat{M} - M) = C[X_{\nu,q}]_i(\hat{M} - M) + D[W_{\nu,q}]_i(\hat{M} - M)$ and by using (19):

$$\begin{aligned} [Y_{\nu,q}]_i (\hat{M} - M) = {} & CA[X_{\nu,q}]_i (\hat{M} - M) \\ & + (CE + D)[W_{\nu,q}]_i \hat{M}, + D[W^{\text{diff}}_{\nu,q}]_i \hat{M}. \end{aligned}$$

Finally, by iterating *Step 2* in the proof of Theorem 4.1, we obtain $[X_{\nu,q}]_i M = (I - A)^{-1} B$, which proves the claim. ■

Proposition 4.3 shows that the absolute error $\hat{G} - G$ is governed by three terms: (i) the difference $[X_{\nu,q}]_i \hat{M} - (I - A)^{-1} B$, which describes the error between any equilibrium state that is compatible with the noisy data $[X_{\nu,q}]_i \hat{M} u$ and the true equilibrium state, described by $(I - A)^{-1} B$; (ii) the quantity $E[W_{\nu,q}]_i \hat{M}$, which describes the error due to noise affecting the model equation; (iii) the quantity $[W^{\text{diff}}_{\nu,q}]_i \hat{M}$, which accounts for the effect of noise on the output equation.

We note that if $\hat{M}$ satisfies (17) and, simultaneously, also $W^{\text{diff}}_{\nu,q} \hat{M} = 0$, and $W_{\nu,q} \hat{M} = 0$, then we have $[X_{\nu,q}]_i \hat{M} = (I - A)^{-1} B$, and thus $\hat{G} - G = 0$. Hence, in this case, we recover the characterization presented in Theorem 4.1.

Similarly to Corollary 4.2, we have the following data-explicit expression for $\hat{G}$, which follows directly from (17).

Corollary 4.4: ***(Data-based Expression of $\hat{G}$)*** Under the same assumptions of Proposition 4.3, matrix $\hat{G}$ computed as

$$\hat{G} = [Y_{\nu,q}]_i \begin{bmatrix} Y^{\text{diff}}_{\nu,q} \\ U_{\nu,q} \end{bmatrix}^\dagger \begin{bmatrix} 0 \\ \mathbb{1}_\nu \otimes I_m \end{bmatrix}, \quad (20)$$

satisfies (18), with $\hat{M} = \begin{bmatrix} Y^{\text{diff}}_{\nu,q} \\ U_{\nu,q} \end{bmatrix}^\dagger \begin{bmatrix} 0 \\ \mathbb{1}_\nu \otimes I_m \end{bmatrix}$. □

Remark 3: ***(Direct vs Indirect Approach to Computing $\hat{G}$)*** The results derived in this section shed some light on the advantages and disadvantages of the direct versus the indirect approach, a matter which is the subject of ongoing research, cf. [43], [44]. In the present context, by direct approach we mean the computation of the steady-state transfer function via algebraic operations on the data. By indirect approach, we mean first performing system identification (namely, deriving matrices (A, B, C, D, E) from historical data [45]) and subsequently using the estimates to compute the steady-state transfer function. We observe the following benefits of the direct approach over the indirect one. The first benefit is computational: (13) and (20) allow us to characterize the system behavior by performing simple algebraic operations, thus avoiding the need to resort to a more-advanced system identification techniques (which often involve two steps: (i) the extraction of the extended observability matrix, possibly after a first step where the impulse response is estimated, to

[4]Notice that the computation of these Hankel matrices requires the availability of a $(T + 1)$-long signal $u_{[0,T]}$, and of a $(T + 2)$-long signal $y_{[0,T+1]}$ (where the additional sample is needed to compute the difference signal). By comparison, the only additional complexity due to the presence of an unknown but constant disturbance consists in collecting one additional sample as compared to the case where the disturbance is known.

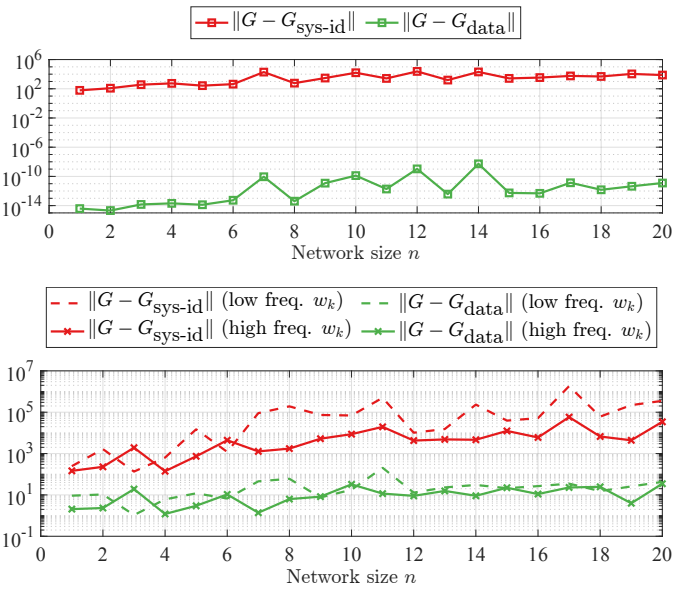


Fig. 1. Montecarlo simulations comparing the accuracy of (20) with that of system identification methods. (Top) illustrates the noiseless case $w_{[0,T]} = 0$ at all times; (Bottom) illustrates the noisy case, where $w_{[0,T]}$ is chosen as a piecewise-constant random signal, and where the frequency of variation is varied over two cases (low frequency and high frequency). System identification was done using a subspace identification algorithm that minimizes the prediction errors to obtain maximum-likelihood estimates (see [45, Sec 7.3]). The curves illustrate the average over 100 experiments, where matrices A and B have been populated with random i.i.d. normal entries and where the modulus of the eigenvalues of A has been chosen in the real interval $(-1, 1)$.

reconstruct the state of the system from input-output data, and (ii) solving a convex optimization problem to determine the matrices (A, B, C, D, E), see [45, Sec 7.3]). Second, Proposition 4.3 provides an explicit error characterization for $\hat{G}$, which is typically not available through system identification methods since error estimates for the matrices (A, B, C) are generally unknown. Third, our numerical simulations suggest that our methods are more accurate in estimating $\hat{G}$ from noisy (and non-noisy) data, cf. Fig. 1. As shown by the top plot, in the absence of noise (i.e., $w_{[0,T]} = 0$), equation (13) allows us to recover the steady-state transfer function up to a numerical error of the order 10^{-14}. In contrast, subspace identification methods compute approximations with errors in the order 10^2. The reason why subspace identification methods have poor performance is that they are designed to compute a set of matrices (A, B, C, D) that minimize the prediction error for the given data, while they fall short of computing a set of matrices (A, B, C, D) approximating those that originated the data. From our simulations, we also observed that subspace identification methods are more precise when the system state can be measured directly, but perform poorly when one measures a low-dimensional output vector (as in our setting); this is because identification methods first reconstruct the internal state of the system using the outputs. Fig. 1(bottom) illustrates that our methods are more accurate than system identification also in the presence of noise. □

V. Synthesis of Stochastic Optimization-based Controllers for Dynamical Systems

In this section, we build upon the characterizations in Section IV and derive a control method that addresses challenge (C2). The basic idea to overcome the lack of knowledge of $\mathcal{W}_k$ is using samples of the system output y_k.

We begin by illustrating the controller synthesis technique. To this end, we notice that a gradient-descent iteration to determine a solution of the optimization problem (4) reads as:

$$u_{k+1} = u_k - \eta \mathbb{E}_{w_k}[\nabla_u \phi(u_k, Gu_k + Hw_k) + G^{\mathsf{T}} \nabla_y \phi(u_k, Gu_k + Hw_k)], \tag{21}$$

where $\eta \in \mathbb{R}_{>0}$ is a tunable controller gain. Unfortunately, the update (21) suffers from the following three main limitations: (i) evaluations of the right-hand side of (21) require knowledge of the transfer function G, which is known only approximately through $\hat{G}$, (ii) the update requires evaluations of the gradient functions at the points $Gu_k + Hw_k$, which are unknown when G and H are unknown (notice that $Gu_k + Hw_k$ is not only unknown but also unmeasurable, since $y_k = Gu_k + Hw_k$ does not hold unless the plant has infinitely-fast dynamics), and (iii) computing the expectation in (21) requires the knowledge of the distributions $\mathcal{W}_k$, for all $k \in \mathbb{Z}_{>0}$, which are also unknown. In order to overcome limitation (i), we replace G with its approximation (20) computed from data; in order to deal with limitation (ii), we replace the steady-state map $Gu_k + Hw_k$ with instantaneous samples of the output y_k (thus making the algorithm *online* [23]); to cope with limitation (iii), we replace exact gradient evaluations with instantaneous samples (thus making the algorithm *stochastic*). This gives rise to the following data-enabled stochastic gradient-descent controller:

$$x_{k+1} = Ax_k + Bu_k + Ew_k, \quad y_k = Cx_k + Dw_k, \tag{22}$$
$$u_{k+1} = u_k - \eta\Big(\nabla_u \phi(u_k, y_k) + ([Y_{\nu,q}]_i \begin{bmatrix} Y^{\text{diff}}_{\nu,q} \\ U_{\nu,q} \end{bmatrix}^{\dagger} \begin{bmatrix} 0 \\ \mathbb{1}_\nu \otimes I_m \end{bmatrix})^{\mathsf{T}} \nabla_y \phi(u_k, y_k)\Big),$$

where the characterization (20) has been used to replace G. Notice that the controller update(22) can be evaluated by only having access to the gradients of the cost functions of (4) and to (possibly noisy) historical data generated by the system.

Remark 4: ***(Extensions to Constrained Optimization Problems and Time-Varying Costs)*** The proposed framework can be extended to control problems (4) that include constraints and/or time-varying costs. When the optimization problem (4) includes constrains of the form $\bar{u} \in \mathcal{U}$, where $\mathcal{U} \subseteq \mathbb{R}^m$ is closed and convex, then (22) can be modified as:

$$u_{k+1} = \Pi_{\mathcal{U}}\Big[u_k - \eta\Big(\nabla_u \phi(u_k, y_k) + \Big([Y_{\nu,q}]_i \begin{bmatrix} Y^{\text{diff}}_{\nu,q} \\ U_{\nu,q} \end{bmatrix}^{\dagger} \begin{bmatrix} 0 \\ \mathbb{1}_\nu \otimes I_m \end{bmatrix}\Big)^{\mathsf{T}} \nabla_y \phi(u_k, y_k)\Big)\Big], \tag{23}$$

where $\Pi_{\mathcal{U}} : \mathbb{R}^m \to \mathcal{U}$ denotes the orthogonal projection onto $\mathcal{U}$, namely, $\Pi_{\mathcal{U}}(z) := \arg\min_{u \in \mathcal{U}} \|u - z\|$, for any $z \in \mathbb{R}^m$. In this case, by using the non-expansiveness property of the projection operator [40], all the conclusions drawn in the remainder of this paper hold unchanged. When the cost of

(4) is time-dependent, i.e., $\phi(u,y)$ is generalized to $\phi(u,y,k)$, all the results derived in the remainder also hold unchanged, provided that Assumption 2 holds uniformly in time. □

Remark 5: ***(Extensions to non-stable systems)*** It is often of interest to account for cases where (2) is not a stable plant. In general, marginally stable or unstable systems are more challenging to optimize since in these cases (4) can no longer be written as an unconstrained optimization problem. To account for these cases, one could decompose the control input u_k as follows $u_k = u_k^{\mathrm{s}} + u_k^{\mathrm{o}}$, where u_k^{s} is used for stabilization and u_k^{o} is used for optimization. With this separation, the techniques in [14] can be used to design an output-feedback controller applied through u_k^{s}, while the methods in our work can be directly applied to design u_k^{o}. □

VI. Notion of Stable Optimizer

As anticipated in (C3), because of the direct dependence between the random variable z_k and the decision variable $\bar{u}$ in (5), the control algorithm (22) might be incapable of determining the exact solution of the optimization problem[5] (5). For this reason, in what follows we focus on a relaxed notion of optimizer of (5), as formalized next.

Definition 6.1: ***(Stable Optimizer)*** The vector $u_k^{\mathrm{so}} \in \mathbb{R}^m$ is a stable optimizer of (5) at time $k \in \mathbb{Z}_{>0}$ if:

$$u_k^{\mathrm{so}} = \arg\min_{\bar{u}} \quad \mathbb{E}_{\mathcal{Z}_k(u_k^{\mathrm{so}})}\Big[\phi(\bar{u}, \hat{G}\bar{u} + z_k)\Big]. \tag{24}$$

Accordingly, we let $x_k^{\mathrm{so}} := (I-A)^{-1}Bu_k^{\mathrm{so}} + (I-A)^{-1}Ew_k$. □

In words, u_k^{so} is a stable optimizer if it solves the optimization problem that originates by fixing the distribution of z_k to $\mathcal{Z}_k(u_k^{\mathrm{so}})$. Convergence to a stable optimizer is desirable because it guarantees that, when u_k^{so} is applied as an input to (2), the resulting cost is optimal for the distribution induced by the random variables. In general, stable optimizers $(u_k^{\mathrm{so}}, x_k^{\mathrm{so}})$ do not coincide with the true optimizer (u_k^*, x_k^*) of (5), and existence and uniqueness are guaranteed under suitable technical assumptions (see Theorem 7.1). We illustrate the distinction between the two concepts in the following example.

Example 1: ***(Stable Optimizers vs True Optimizers)*** Consider an instance of (4) for a scalar dynamical system (i.e., $n = m = p = r = 1$), so that $G, H \in \mathbb{R}$, and $\phi(u,y) = u^2 + y$. In this case, the objective of (4) can be specified in closed form as $\mathbb{E}[\phi(\bar{u}, G\bar{u} + Hw_k)] = \bar{u}^2 + G\bar{u} + H\mathbb{E}[w_k]$, and thus the optimizer of (4) is $u_k^* = -\frac{G}{2}$. To determine a stable optimizer, let $\hat{G}$ be an approximation of G, define $z_k := (G - \hat{G})\bar{u} + Hw_k$, and rewrite the objective as $\mathbb{E}[\phi(\bar{u}, G\bar{u} + Hw_k)] = \mathbb{E}[\phi(\bar{u}, \hat{G}\bar{u} + z_k)] = \mathbb{E}[\bar{u}^2 + \hat{G}\bar{u} + z_k]$. Next, notice that $\nabla_{\bar{u}}\phi(\bar{u}, \hat{G}\bar{u} + z_k) = 2\bar{u} + \hat{G}$, and thus u_k^{so} satisfies $\mathbb{E}[2\bar{u} + \hat{G}] = 0$, which implies $u_k^{\mathrm{so}} = -\frac{\hat{G}}{2}$. □

While true optimizer and stable optimizers do not coincide in general, an explicit error bound can be derived under suitable smoothness assumptions, as shown next.

[5]A similar behavior also originates when gradient-descent is applied to static optimization problems (i.e. without plants), as outlined in [31]–[33].

Proposition 6.2: ***(Optimizer Gap)*** Let Assumption 2 hold, let u_k^* be as in (5), and let u_k^{so} be a stable optimizer. Then,

$$\|u_k^* - u_k^{\mathrm{so}}\| \leq \frac{2\ell\|G - \hat{G}\|}{\mu\sigma_{\min}^2(\hat{G})}, \tag{25}$$

where $\sigma_{\min}^2(\hat{G})$ denotes the smallest singular value of $\hat{G}$.

Proof: By recalling that $\bar{z} = (G - \hat{G})\bar{u} + w_k$, a direct application of Theorem 2.4 yields:

$$W_1(\mathcal{Z}_k(u), \mathcal{Z}_k(u')) \leq \|G - \hat{G}\|\|u - u'\|, \tag{26}$$

for any $u, u' \in \mathbb{R}^m$. Next, we denote in compact form $f(u,z) := \phi(u, \hat{G}u + z)$. By recalling the definition of u_k^* and of u_k^{so}, we have that $\mathbb{E}_{\mathcal{Z}(u_k^*)} f(u_k^*, z) \leq \mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})} f(u_k^{\mathrm{so}}, z)$, which implies:

$$\begin{aligned}\mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[f(u_k^*, z)\right] - \mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[f(u_k^{\mathrm{so}}, z)\right] \\ \leq \mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[f(u_k^*, z)\right] - \mathbb{E}_{\mathcal{Z}(u_k^*)}\left[f(u_k^*)\right].\end{aligned} \tag{27}$$

First, we upper bound the right-hand side of (27). To this aim, by combining (26) with Assumption 2(a) and by application of Lemma 2.5 we have:

$$\mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[f(u_k^*, z)\right] - \mathbb{E}_{\mathcal{Z}(u_k^*)}\left[f(u_k^*, z)\right] \leq \ell\|G - \hat{G}\|\|u_k^* - u_k^{\mathrm{so}}\|. \tag{28}$$

Second, we lower-bound the left-hand side of (27). To this aim, we note that Assumption 2(c) implies: $f(u_k^*, z) \geq f(u_k^{\mathrm{so}}, z) + \nabla_u f(u_k^{\mathrm{so}}, z)^{\mathsf{T}}(u_k^* - u_k^{\mathrm{so}}) + \frac{\mu\sigma_{\min}^2(\hat{G})}{2}\|u_k^* - u_k^{\mathrm{so}}\|^2$ for all z. Moreover, since u_k^{so} is a stable optimizer, it satisfies the following variational inequality: $\mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}[\nabla_u f(u_k^{\mathrm{so}}, z)^{\mathsf{T}}(u' - u_k^{\mathrm{so}})] \geq 0$ for all $u' \in \mathbb{R}^m$. By combining the above two conditions:

$$\mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[f(u_k^*, z)\right] - \mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[f(u_k^{\mathrm{so}}, z)\right] \geq \frac{\mu\sigma_{\min}^2(\hat{G})}{2}\|u_k^{\mathrm{so}} - u_k^*\|^2. \tag{29}$$

Finally, the claim follows by combining (28) with (29). ■

Proposition 6.2 quantifies the gap between a stable optimizer and the true optimizer of (5). The bound shows that the gap grows linearly with the absolute error $\|G - \hat{G}\|$ and with the Lipschitz constant ℓ, and is inversely proportional to the smallest singular value of $\hat{G}$ and to the strong-convexity constant μ. Finally, we note that when G is known exactly (i.e., $\hat{G} = G$), then (25) yields $u_k^* = u_k^{\mathrm{so}}$. Indeed, in this case (5) does not feature distributions that are decision-dependent.

VII. Controller Analysis in the Presence of Time-Varying Disturbances

Having presented solutions to the challenges (C1)-(C3), we are now ready to characterize the transient performance of the controller proposed in (22) with respect to a stable optimizer. More generally, in this section, we focus on stochastic gradient descent controllers with an inexact gradient of the form:

$$\begin{aligned}&x_{k+1} = Ax_k + Bu_k + Ew_k, \quad y_k = Cx_k + Dw_k,\\ &u_{k+1} = u_k - \eta\Big(\nabla_u\phi(u_k, y_k) + \hat{G}^{\mathsf{T}}\nabla_y\phi(u_k, y_k)\Big),\end{aligned} \tag{30}$$

where $\hat{G}$ denotes a deterministic approximation of G.

Remark 6: ***(Generality of the Analysis)*** While Section IV presents a way to compute $\hat{G}$ from data, the results presented here hold for any deterministic $\hat{G}$ (computed by using other methods such as, e.g., system identification [45]), provided that a deterministic error bound for $\|G-\hat{G}\|$ is available. □

For the subsequent analysis, we let

$$e_k := \nabla_u \phi(u_k, y_k) + \hat{G}^{\mathsf{T}} \nabla_y \phi(u_k, y_k) - \mathbb{E}_{y_k}[\nabla_u \phi(u_k, y_k) + \hat{G}^{\mathsf{T}} \nabla_y \phi(u_k, y_k)], \quad (31)$$

denote the error that originates from approximating the true gradient (obtained by computing the expectation) with a single-point gradient[6] (obtained by sampling the random variables through the output of the dynamical system).

Theorem 7.1: ***(Tracking of Time-Varying Stable Optimizer)*** Let Assumption 2 be satisfied. If a stable optimizer $\xi_k^{\mathrm{so}} := (x_k^{\mathrm{so}}, u_k^{\mathrm{so}})$ exists, then for any $k \in \mathbb{Z}_{>0}$ the solutions $\xi_k := (x_k, u_k)$ of (30) satisfy:

$$\mathbb{E}[\|\xi_{k+1} - \xi_{k+1}^{\mathrm{so}}\|] \leq \beta_1 \mathbb{E}\left[\|u_k - u_k^{\mathrm{so}}\|\right] + \beta_2 \mathbb{E}\left[\|x_k - x_k^{\mathrm{so}}\|\right] + \gamma_1 \mathbb{E}[\|e_k\|] + \gamma_2 \|u_{k+1}^{\mathrm{so}} - u_k^{\mathrm{so}}\| + \gamma_3 \mathbb{E}[\sup_{t \in \mathbb{Z}_{>0}} \|x_{t+1}^{\mathrm{so}} - x_t^{\mathrm{so}}\|], \quad (32)$$

where e_k defined in (31) and, for any $\kappa \in (0,1)$,

$$\beta_1 = \sqrt{1-\eta\mu} + \eta \hat{\ell}^{\nabla} \|G - \hat{G}\|, \quad \hat{\ell}^{\nabla} := \ell_u^{\nabla} + \|\hat{G}\| \ell_y^{\nabla},$$

$$\beta_2 = \sqrt{\frac{\bar{\lambda}(P)}{\underline{\lambda}(P)} \left(1 - (1-\kappa)\frac{\underline{\lambda}(Q)}{\bar{\lambda}(P)}\right)} + \eta \hat{\ell}^{\nabla} \|C\|,$$

$$\gamma_1 = \eta, \quad \gamma_2 = 1, \quad \gamma_3 = \max\{\sqrt{\frac{2\bar{\lambda}(P)}{\kappa \underline{\lambda}(Q)}}, \frac{4\|A^{\mathsf{T}} P\|}{\kappa \underline{\lambda}(Q)}\}.$$

Moreover, if $\beta_1, \beta_2 < 1$, then $(x_k^{\mathrm{so}}, u_k^{\mathrm{so}})$ exists and is unique. □

Proof: The proof is organized into five main steps.

(1 – Change of Variables and Contraction Bound) Define the change of variables $\tilde{x}_k := x_k - x_k^{\mathrm{so}} = x_k - (I-A)^{-1} B u_k^{\mathrm{so}} - (I-A)^{-1} E w_k$. Accordingly, (30) read as:

$$\tilde{x}_{k+1} = A\tilde{x}_k + (x_k^{\mathrm{so}} - x_{k+1}^{\mathrm{so}}),$$
$$u_{k+1} = u_k - \eta(\nabla_u \phi(u_k, C\tilde{x}_k + \hat{G}u_k + z_k) + \hat{G}^{\mathsf{T}} \nabla_u \phi(u_k, C\tilde{x}_k + \hat{G}u_k + z_k)).$$

Next, we introduce the following compact notation to denote the algorithmic updates (30) for all $\theta \in \mathbb{R}^m$, $u \in \mathbb{R}^m$, $x \in \mathbb{R}^n$:

$$F_\theta(u,x) := \mathop{\mathbb{E}}_{z_k \sim \mathcal{Z}(\theta)}[\nabla_u \phi(u, Cx + \hat{G}u + z_k) + \hat{G}^{\mathsf{T}} \nabla_y \phi(u, Cx + \hat{G}u + z_k)], \quad (33)$$

$$\hat{F}(u,x) := \nabla_u \phi(u, Cx + \hat{G}u + z_k) + \hat{G}^{\mathsf{T}} \nabla_y \phi(u, Cx + \hat{G}u + z_k),$$

$$\mathcal{C}_\theta(u,x) := u - \eta F_\theta(u,x), \quad \hat{\mathcal{C}}(u,x) := u - \eta \hat{F}(u,x).$$

Accordingly, the left hand side of (32) satisfies:

$$\mathbb{E}[\|\xi_{k+1} - \xi_{k+1}^{\mathrm{so}}\|] \leq \mathbb{E}[\|u_{k+1} - u_{k+1}^{\mathrm{so}}\|] + \mathbb{E}[\|\tilde{x}_{k+1}\|] \quad (34)$$
$$\leq \mathbb{E}[\|u_{k+1} - u_k^{\mathrm{so}}\|] + \|u_{k+1}^{\mathrm{so}} - u_k^{\mathrm{so}}\| + \mathbb{E}[\|\tilde{x}_{k+1}\|],$$

[6]Notice that in (31) both the single-point and exact gradients are evaluated at the same point (u_k, y_k). Thus e_k does *not* account for errors due to evaluating gradients at y_k instead of $Gu_k + Hw_k$ (cf. Remark 12).

where we used $\mathbb{E}[\|u_{k+1}^{\mathrm{so}} - u_k^{\mathrm{so}}\|] = \|u_{k+1}^{\mathrm{so}} - u_k^{\mathrm{so}}\|$ since stable optimizers are deterministic quantities. Moreover, notice that:

$$\begin{aligned}\mathbb{E}[\|u_{k+1} - u_k^{\mathrm{so}}\|] &= \mathbb{E}[\|\hat{\mathcal{C}}(u_k, \tilde{x}_k) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k^{\mathrm{so}}, 0)\|] \\ &\leq \mathbb{E}[\|e_k\|] + \|\mathcal{C}_{u_k}(u_k, \tilde{x}_k) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k, \tilde{x}_k)\| \\ &\quad + \|\mathcal{C}_{u_k^{\mathrm{so}}}(u_k, \tilde{x}_k) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k, 0)\| \\ &\quad + \|\mathcal{C}_{u_k^{\mathrm{so}}}(u_k, 0) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k^{\mathrm{so}}, 0)\|. \end{aligned} \quad (35)$$

where we used $\hat{\mathcal{C}}(u_k, \tilde{x}) - \mathcal{C}_{u_k}(u_k, \tilde{x}) = e_k$ and we remark that the last three terms are deterministic quantities. Linear convergence of (30) is a direct consequence of four independent properties, namely calmness to distributional shifts, ease with respect to system dynamics, contraction at the equilibrium, and contraction of the dynamical system, which we prove next.

(2 – Calmness With Respect to Distributional Shifts) We will show: $\|\mathcal{C}_{u_k}(u_k, \tilde{x}_k) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k, \tilde{x}_k)\| \leq \eta \hat{\ell}^{\nabla} \|G - \hat{G}\| \|u_k - u_k^{\mathrm{so}}\|$. Indeed, the following estimate holds:

$$\begin{aligned}\|\mathcal{C}_{u_k}(u_k, \tilde{x}_k) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k, \tilde{x}_k)\| &\leq \eta \hat{\ell}^{\nabla} W_1(\mathcal{Z}(u_k), \mathcal{Z}(u_k^{\mathrm{so}})) \\ &\leq \eta \hat{\ell}^{\nabla} \|G - \hat{G}\| \|u_k - u_k^{\mathrm{so}}\|, \end{aligned}$$

where the first inequality follows by expanding (33) and by using Lemma 2.5, and the second inequality follows from (26).

(3 – Ease With Respect to the System Dynamics) We will show that $\|\mathcal{C}_{u_k^{\mathrm{so}}}(u_k, \tilde{x}_k) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k, 0)\| \leq \eta \hat{\ell}^{\nabla} \|C\| \|\tilde{x}_k\|$. By using Assumption 2(b):

$$\begin{aligned}&\|\mathcal{C}_{u_k^{\mathrm{so}}}(u_k, \tilde{x}_k) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k, 0)\| \\ &\leq \eta \|\mathbb{E}_{\mathcal{Z}_k(u_k^{\mathrm{so}})}\left[\nabla_u \phi(u_k, C\tilde{x}_k + \hat{G}u_k + z_k) - \nabla_u \phi(u_k, \hat{G}u_k + z_k)\right]\| \\ &\quad + \eta \|\hat{G}^{\mathsf{T}} \mathbb{E}_{\mathcal{Z}_k(u_k^{\mathrm{so}})}\left[\nabla_y \phi(u_k, C\tilde{x}_k + \hat{G}u_k + z_k) - \nabla_y \phi(u_k, \hat{G}u_k + z_k)\right]\| \\ &\leq \eta \hat{\ell}^{\nabla} \|C\| \|\tilde{x}_k\|, \end{aligned}$$

which proves the claimed estimate.

(4 – Contraction at the Equilibrium) We will show that $\|\mathcal{C}_{u_k^{\mathrm{so}}}(u_k, 0) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k^{\mathrm{so}}, 0)\| \leq \sqrt{1-\eta\mu} \|u_k - u_k^{\mathrm{so}}\|$. This fact follows directly from [37, Thm 3.12], and we provide a short proof for completeness. By substituting (33):

$$\begin{aligned}&\|\mathcal{C}_{u_k^{\mathrm{so}}}(u_k, 0) - \mathcal{C}_{u_k^{\mathrm{so}}}(u_k^{\mathrm{so}}, 0)\|^2 = \|u_k - \eta F_{u_k^{\mathrm{so}}}(u_k, 0) - u_k^{\mathrm{so}}\|^2 \\ &= \|u_k - u_k^{\mathrm{so}}\|^2 - 2\eta F_{u_k^{\mathrm{so}}}(u_k, 0)^{\mathsf{T}}(u_k - u_k^{\mathrm{so}}) + \eta^2 \|F_{u_k^{\mathrm{so}}}(u_k, 0)\|^2 \\ &\leq (1-\eta\mu)\|u_k - u_k^{\mathrm{so}}\|^2 - 2\eta\Big(\mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[\phi(u_k, C\hat{u}_k + z_k)\right] \\ &\quad - \mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[\phi(u_k^{\mathrm{so}}, \hat{G}u_k^{\mathrm{so}} + z_k)\right]\Big) + \eta^2 \|F_{u_k^{\mathrm{so}}}(u_k, 0)\|^2 \\ &\leq (1-\eta\mu)\|u_k - u_k^{\mathrm{so}}\|^2 + \alpha\Big(\mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[\phi(u_k, \hat{G}u_k + z_k)\right] \\ &\quad - \mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}\left[\phi(u_k^{\mathrm{so}}, \hat{G}u_k^{\mathrm{so}} + z_k)\right]\Big) \\ &\leq (1-\eta\mu)\|u_k - u_k^{\mathrm{so}}\|^2, \end{aligned}$$

where $\alpha = 2(\eta^2 \hat{\ell}^{\nabla} - \eta)$. Above, the first inequality follows from $\phi(u_k^{\mathrm{so}}, y) - \phi(u_k, y_k) \geq \nabla \phi(u_k^{\mathrm{so}}, y)^{\mathsf{T}}(u_k^{\mathrm{so}} - u_k) + \frac{\mu}{2}\|u_k - u_k^{\mathrm{so}}\|^2$ (see Assumption 2(c)), the second inequality follows from $\|F_{u_k^{\mathrm{so}}}(u_k, 0)\|^2 \leq 2\hat{\ell}^{\nabla} \mathbb{E}_{\mathcal{Z}(u_k^{\mathrm{so}})}[\phi(u_k, \hat{G}u_k + z_k) - \phi(u_k^{\mathrm{so}}, \hat{G}u_k^{\mathrm{so}} + z_k)]$ (see Assumption 2(b)), and the last inequality holds because u_k^{so} is a stable optimizer (see (24)).

(5 – Contraction of the Dynamical System) We will prove the following estimate:

$$\mathbb{E}[\|\tilde{x}_{k+1}\|] \leq \sqrt{\frac{\bar{\lambda}(P)}{\underline{\lambda}(P)}\left(1-\frac{\underline{\lambda}(Q)}{4\bar{\lambda}(P)}\right)}\mathbb{E}[\|\tilde{x}_k\|] \tag{36}$$
$$+\max\{\sqrt{\frac{4\bar{\lambda}(P)}{\underline{\lambda}(Q)}},\frac{4\|A^\mathsf{T}P\|}{\underline{\lambda}(Q)}\}\mathbb{E}[\sup_{t\in\mathbb{Z}_{>0}}\|x^{\text{so}}_{k+1}-x^{\text{so}}_k\|].$$

In what follows, we fix the realization of the disturbance w_k and (with a slight abuse of notation) we denote by $\tilde{x}_k$ the corresponding (deterministic) state of (2) and by x^{so}_k the associated (deterministic) stable optimizer. Let $V(x) := x^\mathsf{T}Px$ and define the set:

$$\Omega := \{x\in\mathbb{R}^n : V(x) \leq \frac{4\bar{\lambda}(P)\|A^\mathsf{T}P\|}{\underline{\lambda}(Q)}\sup_{t\in\mathbb{Z}_{>0}}\|x^{\text{so}}_{t+1}-x^{\text{so}}_t\|$$
$$\text{and } V(x)\leq\bar{\lambda}(P)\sqrt{\frac{4\bar{\lambda}(P)}{\underline{\lambda}(Q)}}\sup_{t\in\mathbb{Z}_{>0}}\|x^{\text{so}}_{t+1}-x^{\text{so}}_t\|\}.$$

We distinguish among two cases.

(5 – Case 1) Suppose $\tilde{x}_k\notin\Omega$. In this case, we have:

$$\begin{aligned}V(\tilde{x}_{k+1})-V(\tilde{x}_k) &\leq -\underline{\lambda}(Q)\|\tilde{x}_k\|^2+\bar{\lambda}(P)\|x^{\text{so}}_{k+1}-x^{\text{so}}_k\|^2\\ &\quad+2\|A^\mathsf{T}P\|\|x^{\text{so}}_{k+1}-x^{\text{so}}_k\|\|\tilde{x}_k\|\\ &\leq -\frac{1}{4}\frac{\underline{\lambda}(Q)}{\bar{\lambda}(P)}V(\tilde{x}_k),\end{aligned}\tag{37}$$

where the last inequality follows since $\tilde{x}_k\notin\Omega$ and by using $V(\tilde{x}_k)\leq\bar{\lambda}(P)\|\tilde{x}_k\|^2$. By using $\underline{\lambda}(P)\|\tilde{x}_k\|^2\leq V(\tilde{x}_k)\leq\bar{\lambda}(P)\|\tilde{x}_k\|^2$, (37) implies the following bound for the state:

$$\|\tilde{x}_{k+1}\|^2\leq\frac{\bar{\lambda}(P)}{\underline{\lambda}(P)}\left(1-\frac{\underline{\lambda}(Q)}{4\bar{\lambda}(P)}\right)\|\tilde{x}_k\|^2.\tag{38}$$

(5 – Case 2) Suppose $\tilde{x}_k\in\Omega$. In this case, we will show that Ω is forward-invariant, i.e., $\tilde{x}_{k+1}\in\Omega$. By contradiction, let $\epsilon>0$ and let k_1 be the first instant such that one of the following conditions is satisfied:

$$V(\tilde{x}_{k_1})>\bar{\lambda}(P)\frac{4\|A^\mathsf{T}P\|}{\underline{\lambda}(Q)}\sup_{t\in\mathbb{Z}_{>0}}\|x^{\text{so}}_{t+1}-x^{\text{so}}_t\|+\epsilon,\text{ or}$$
$$V(\tilde{x}_{k_1})>\bar{\lambda}(P)\sqrt{\frac{4\bar{\lambda}(P)}{\underline{\lambda}(Q)}}\sup_{t\in\mathbb{Z}_{>0}}\|x^{\text{so}}_{t+1}-x^{\text{so}}_t\|+\epsilon,\tag{39}$$

It follows by iterating (37) that $V(\tilde{x}_k)$ is strictly decreasing in a neighborhood of k_1. Accordingly, there must exist $0\leq k_0<k_1$ such that $V(\tilde{x}_{k_0})>V(\tilde{x}_{k_1})$. But this contradicts the assumption that k_1 is the first instant that satisfies (39). So Ω must be forward invariant. By recalling the definition of Ω, when $\tilde{x}_k\in\Omega$:

$$\|\tilde{x}_{k+1}\|\leq\max\{\sqrt{\frac{4\bar{\lambda}(P)}{\underline{\lambda}(Q)}},\frac{4\|A^\mathsf{T}P\|}{\underline{\lambda}(Q)}\}\sup_{t\in\mathbb{Z}_{>0}}\|x^{\text{so}}_{k+1}-x^{\text{so}}_k\|.\tag{40}$$

Finally, the estimate (36) follows by combining (38) and (40) and by taking the expectation on both sides.

To conclude, (32) follows by substituting the estimates derived in the above five steps into (34)–(35). Notice that, the existence and uniqueness of u^{so}_k follows from contractivity and by application of the Banach fixed-point theorem. ■

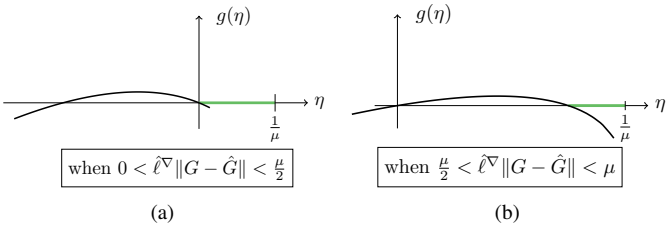


Fig. 2. Roots of $g(\eta)=\sqrt{1-\eta\mu}+\eta\hat{\ell}^\nabla\|G-\hat{G}\|-1$. The green segment illustrates the set of choices of η that guarantee $\beta_1<1$.

Remark 7: ***(Choices of η that Guarantee $\beta_1<1$)*** To guarantee $\beta_1<1$, the following conditions must hold simultaneously:

$$\|G-\hat{G}\|<\frac{\mu}{\hat{\ell}^\nabla},\quad\text{and}\quad\frac{2\|G-\hat{G}\|-\mu}{\|G-\hat{G}\|^2}<\eta\leq\frac{1}{\mu}.\tag{41}$$

Note that for any $\mu\in\mathbb{R}_{>0}$ and $\hat{G}\in\mathbb{R}^{p\times m}$, there exists a nonempty set of choices of η that satisfy the second condition in (41). To see this, notice that $\frac{2\|G-\hat{G}\|-\mu}{\|G-\hat{G}\|^2}\leq\frac{1}{\mu}$ is equivalent to $(\|G-\hat{G}\|-\mu)^2\geq 0$, which is satisfied for any $\mu\in\mathbb{R}_{>0}$ and $\hat{G}\in\mathbb{R}^{p\times m}$. While there always exists a choice of η that guarantees the second condition, the first inequality in (41) outlines a feasibility condition. Namely, when the absolute error $\|G-\hat{G}\|$ is larger than the constant $\mu/\hat{\ell}^\nabla$, then the controller (30) might not guarantee contractivity.

To derive (41), notice that $\sqrt{1-\eta\mu}$ admits a real-valued solution if an only if $\eta\leq 1/\mu$. Moreover, the equation $\sqrt{1-\eta\mu}+\eta\hat{\ell}^\nabla\|G-\hat{G}\|=1$ yields the solution $\eta=\eta_1:=0$ and $\eta=\eta_2:=(2\|G-\hat{G}\|-\mu)/\|G-\hat{G}\|^2$, where η_2 is real only if $\hat{\ell}^\nabla\|G-\hat{G}\|\leq\mu$. Accordingly, $\beta_1<1$ when $\eta<\eta_1$ or $\eta>\eta_2$ (see Fig. 2), which yields (41). □

Remark 8: ***(Choices of η that Guarantee $\beta_2<1$)*** To guarantee $\beta_2<1$ the controller η must be chosen as:

$$\eta<\frac{1}{\hat{\ell}^\nabla\|C\|}\left(1-\sqrt{\frac{\bar{\lambda}(P)}{\underline{\lambda}(P)}\left(1-(1-\kappa)\frac{\underline{\lambda}(Q)}{\bar{\lambda}(P)}\right)}\right).\tag{42}$$

Notice that the quantity $\frac{\bar{\lambda}(P)}{\underline{\lambda}(P)}\left(1-(1-\kappa)\frac{\underline{\lambda}(Q)}{\bar{\lambda}(P)}\right)$ is always non-negative since $\underline{\lambda}(Q)/\bar{\lambda}(P)<1$ and it strictly smaller that 1 if the open-loop dynamics (2) are contractive. □

Theorem 7.1 provides a sufficient condition to guarantee that the controlled dynamics (30) converge to a stable optimizer (x^*_k,u^*_k) (up to an asymptotic error that depends on the time-variability of the optimizer and on the sampling error). This result, combined with Proposition 6.2, allows us to conclude convergence to a small neighborhood of the desired optimizer (x^*_k,u^*_k). The estimate (32) upper bounds the tracking error at time $k+1$ (i.e., $\mathbb{E}[\|\xi_{k+1}-\xi^{\text{so}}_{k+1}\|]$) as a function of three main terms: (i) the tracking error at time k (characterized by $\mathbb{E}[\|u_k-u^{\text{so}}_k\|]$ and $\mathbb{E}[\|x_k-x^{\text{so}}_k\|]$), (ii) the error due to sampling (i.e., $\mathbb{E}[\|e_k\|]$), and (iii) the temporal shift of the optimizer (namely, the temporal shift of the input

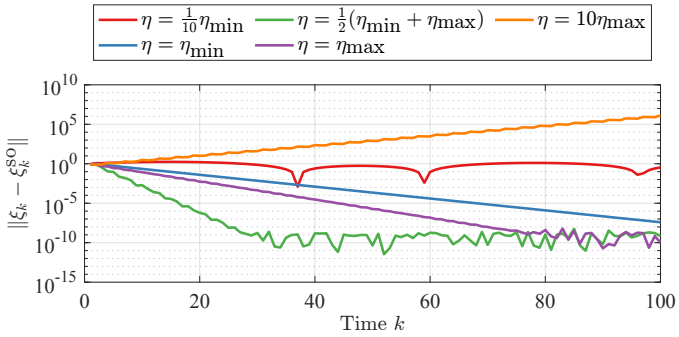


Fig. 3. Numerical simulations illustrating the rate of convergence of (30) for different choices of η. We have conducted the simulations on a system with matrices (A, B, C, D, E) having random entries with $n = 10, m = 5, p = 3, r = 5$. $\eta_{\min} := \frac{2\|G-\hat{G}\|-\mu}{\|G-\hat{G}\|^2}$, $\eta_{\max} := \frac{1}{\mu}$ (see (41)). In all the simulations, the parameters satisfy $\|G - \hat{G}\| < \mu/\hat{\ell}^{\nabla}$ (as in (41)) and (42).

optimizer $\|u^{so}_{k+1} - u^{so}_k\|$ and the worst-case temporal shift in the state optimizer $\mathbb{E}[\sup_{t\in\mathbb{Z}_{>0}} \|x^{so}_{t+1} - x^{so}_t\|]$).

According to Theorem 7.1, the rate of contraction of (30) (characterized by the constants β_1 and β_2) depends on the various parameters of the optimization problem as well as of the dynamical system: it increases with the square root of η, μ, and of the ratio $\underline{\lambda}(Q)/\bar{\lambda}(P)$ (that characterizes the rate of convergence of the open-loop plant (2)), and it is inversely proportional to η, $\hat{\ell}^{\nabla}$, $\|G - \hat{G}\|$, and $\|C\|$. Some important comments on the choice of η are in order. First, as discussed in Remark 7, when the distributional shifts originated by the controller update are small (i.e., $\hat{\ell}^{\nabla}\|G - \hat{G}\| < \mu/2$), then a sufficiently-slow controller (i.e, $\eta \leq 1/\mu$) guarantees contraction in (32). On the other hand, when the distributional shifts originated by the controller updates are large (i.e., $\hat{\ell}^{\nabla}\|G - \hat{G}\| > \mu/2$), then there is a lower bound on the required controller gain to guarantee contractivity (namely, $\eta > (2\|G-\hat{G}\|-\mu)/\|G-\hat{G}\|^2$). This fact can be interpreted by noting that a sufficiently-large controller gain guarantees that deviations introduced by shifts in the distribution (i.e. the term $\eta\hat{\ell}^{\nabla}\|G-\hat{G}\|$) are dominated by the the algorithm contractivity towards the optimizer (i.e., the term $\sqrt{1-\eta\mu}$). See the proof of Theorem 7.1, steps 2 and 4. We note that this fact is in contrast with standard conditions for convergence of gradient-descent algorithms without decision-dependent distributions (see e.g. [37]), where arbitrarily-small choices of the controller gain always guarantee contractivity of the updates.

Finally, we illustrate the rate of convergence of (30) via simulations in Fig. 3. The simulations illustrate that choices of η in the interval (41) guarantee contraction of the tracking error and that (30) has exponential rate of convergence (equivalently, linear in log scale), as formally characterized in Theorem 7.1.

Remark 9: ***(Parameters Needed for Controller Design)*** We notice that even when the system (2) is unknown and the interval (41) cannot be computed exactly, the upper bound $\eta = 1/\mu$ can be computed without knowledge of the system (since it depends only on the parameters of the optimization), and thus it can be used as an initial estimate for the controller stepsize; on the other hand, the lower bound for η in (41) may be harder to estimate since it requires an estimation of $\|G - \hat{G}\|$. Moreover, to satisfy the first inequality in (41), one needs to guarantee that $\|G - \hat{G}\| < \frac{\mu}{\hat{\ell}^{\nabla}}$. In this case, since the strong convexity constant μ and Lipschitz constant $\hat{\ell}^{\nabla}$ are often free design parameters (that can be tuned by changing the cost function), computing the exact error $G - \hat{G}$ is also not necessary and thus precise knowledge of (A, B, C) is not needed, provided that $\|G - \hat{G}\|$ is bounded. □

Remark 10: ***Estimates for the temporal shift of the stable optimizers*** An estimate for the temporal shift of stable optimizers (which multiplies the quantities γ_2 and γ_3 in (32)) can be obtained by leveraging the bound in Proposition 6.2, as discussed next. By using the triangle inequality:

$$\begin{aligned}\|u^{so}_{k+1} - u^{so}_k\| &= \|u^*_{k+1} - u^*_k + u^{so}_{k+1} - u^*_{k+1} - u^{so}_k + u^*_k\| \\ &\leq \|u^*_{k+1} - u^*_k\| + \|u^{so}_{k+1} - u^*_{k+1}\| + \|u^{so}_k - u^*_k\| \\ &\leq \|u^*_{k+1} - u^*_k\| + \frac{4\ell\|G - \hat{G}\|}{\mu\sigma^2_{\min}(\hat{G})},\end{aligned}$$

which shows how an explicit bound for $\|u^{so}_{k+1} - u^{so}_k\|$ can be obtained as a function of the temporal shift of the true optimizer $\|u^*_{k+1} - u^*_k\|$. Notice that, in the online optimization literature, $\|u^*_{k+1} - u^*_k\|$ is typically referred to as "path length". Moreover, when u^*_k can be expressed as a function of the disturbance w_k, the path length $\|u^*_{k+1} - u^*_k\|$ can be bounded with a term that captures the temporal variability of the disturbance w_k. However, for cases where u^*_k cannot be made explicit in terms of w_k, tracking bounds in the online optimization literature are expressed in terms of $\|u^*_{k+1} - u^*_k\|$ (see, e.g., [46], [47]). A similar bound holds for $\|x^{so}_{k+1} - x^{so}_k\|$, which can be derived by iterating the reasoning above. □

Remark 11: ***(Common Assumptions That Guarantee Bounded Gradient Error)*** Theorem 7.1 implicitly assumes that the gradient error $\mathbb{E}[\|e_k\|]$ is bounded. Such an assumption is commonly adopted in the literature (see e.g. [48] for a thorough discussion). Commonly-adopted assumptions that guarantee boundedness of the gradient error include uniform boundedness assumptions of the form:

$$\mathbb{E}[\|e_k\|] < \sigma, \text{ for all } k \in \mathbb{Z}_{>0},$$

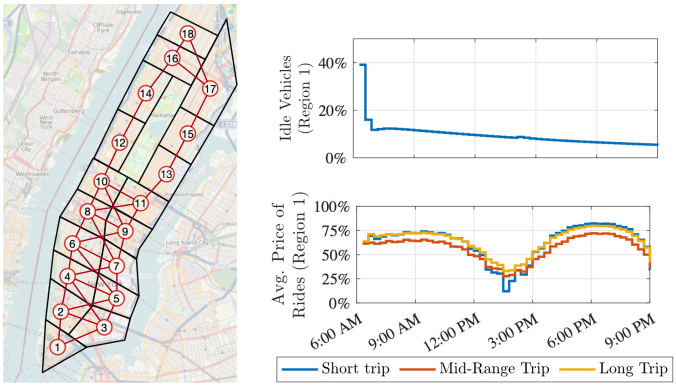
where $\sigma \in \mathbb{R}_{>0}$, or bounded variance assumptions of the form:

$$\mathbb{E}[\|e_k\|^2] \leq \|\mathbb{E}_{y_k}[\nabla_u \phi(u_k, y_k) + \hat{G}^{\mathsf{T}} \nabla_y \phi(u_k, y_k)]\|^2 + \bar{\sigma}^2,$$

for some $\bar{\sigma} \in \mathbb{R}_{\geq 0}$. We also notice that due to unbiasedness, uniform boundedness and bounded variance assumptions are equivalent (see, e.g., [48]). □

Remark 12: ***(Error Due To Sampling vs Error Due To Approximate Evaluation)*** Notice that the term e_k accounts only for the error that is due to the use of a one-sample evaluation of the gradient (compared to the actual gradient obtained by computing the expectation), while it does not include the error due to evaluating the gradient at the instantaneous system output (i.e., at y_k) instead of at the steady-state output (i.e., at $Gu_k + Hw_k$). We remark that the latter is instead accounted through the constant β_2 (see step *(3 – Ease With Respect to the System Dynamics)* in the proof of Theorem 7.1). □

Remark 13: ***(Cases where $\mathcal{W}_k$ is Known)*** In case the distribution $\mathcal{W}_k$ is known for all $k \in \mathbb{Z}_{>0}$, then the true gradient can be computed exactly at each iteration, and thus



(a) (b)

Fig. 4. (a) Case study: New York City, partitioned into 18 regions of ride requests. (b) Number of idle vehicles and cost of trips for region 1. Short trip refers to trips from 1 to 4, mid-range trip refers from 1 to 10, long trip refers from 1 to 16. See caption of Fig. 6 for detailed experiment description.

(32) still holds by letting $e_k = 0$ for all $k \in \mathbb{Z}_{>0}$. In such a case, the obtained contraction bound with respect to the stable optimizer is still relevant as the optimization problem remains decision-dependent (since the steady-state transfer function G remains known only approximately). □

VIII. Application to Ride-Service Scheduling

We illustrate here the versatility and performance of the proposed controller synthesis approach in an application scenario. A ride service provider (RSP), such as Uber, Lyft, or DiDi, seeks to maximize its profit by dispatching the vehicles in its fleet to serve ride requests from its customers. We model the area of interest using a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node in $\mathcal{V}$ represents a region (e.g., a block or a district of a city) and an edge $(i, j) \in \mathcal{E}$ allows rides from node i to node j. As a case study, we consider Manhattan, NY, and, similarly to [49], we divide the area into $n = 18$ region as in Fig. 4(a). We assume that time is slotted and each slot has a duration $\Delta = 5$ min. We let $\delta_k^{ij} \in \mathbb{R}_{\geq 0}$ be the demand of rides from region $i \in \mathcal{V}$ to region $j \in \mathcal{V}$ at time $k \in \mathbb{Z}_{>0}$. We denote by p_{ij}^k the price of rides, decided by the RSP, from region i to region j at time k. We account for elasticity of the demand, whereby customers can decide to accept or decline rides after observing the price set by the RSP, and leave the system when prices are higher than their maximum willingness to pay. We model the elasticity of the demand as follows:

$$d_k^{ij} = \delta_k^{ij}\left(1 - \theta^{ij}\frac{p_k^{ij}}{p_{\max}^{ij}}\right). \tag{43}$$

Here, d_k^{ij} denotes the accepted demand (after customers have observed the set price) from region i to j at time k, $\theta^{ij} \in [0, 1]$ is a parameter that characterizes the steepness of elasticity, and $p_{ij}^{\max} \in \mathbb{R}_{>0}$ is an upper limit on prices from i to j.

We let $x_k^i \in \mathbb{R}_{\geq 0}$ denote the idle-vehicle occupancy (i.e., the number of unoccupied vehicles, normalized by the fleet size) of fleet vehicles in region i at time k. We assume that drivers of unoccupied vehicles naturally rebalance the fleet, namely, they travel from regions with a high occupancy of (fleet) vehicles to regions with a lower occupancy in order to maximize their profit. We denote by $a_{ij} \in \mathbb{R}_{\geq 0}$ the fraction of unoccupied vehicles that travel from i to j at every time step. Travel times are non-negligible and can vary over time, we model them by using Boolean variables:

$$\sigma_k^{ij,\tau} = \begin{cases} 1, & \text{if travel time from } i \text{ to } j \text{ at time } k \text{ is } \tau \text{ slots,} \\ 0, & \text{otherwise,} \end{cases}$$

defined for all $i, j \in \mathcal{V}$ and $k, \tau \in \mathbb{Z}_{>0}$. In what follows, we assume that $\sigma_k^{ij,0} = 0$ for all $i, j \in \mathcal{V}, k \in \mathbb{Z}_{>0}$, so vehicles take at least one time slot to travel between any pair of nodes; moreover, we let $\sigma_k^{ij,\tau} = 0$ for all $\tau > T \in \mathbb{Z}_{>0}$, so that T is the maximum travel time in the network. Accordingly, the occupancy of idle vehicles in each region i satisfies:

$$\begin{aligned} x_{k+1}^i = x_k^i - \sum_{j\in\mathcal{V}} a_{ij}x_k^i + \sum_{j\in\mathcal{V}} a_{ji}x_k^j \\ - \sum_{j\in\mathcal{V}} d_k^{ij} + \underbrace{\sum_{j\in\mathcal{V}}\sum_{\tau=k-T}^{k-1} \sigma_\tau^{ji,k-\tau} d_\tau^{ji} + g_k^i}_{:=w_k^i}, \end{aligned} \tag{44}$$

In (44), the quantity $-\sum_{j\in\mathcal{V}} a_{ij}x_k^i$ accounts for the vehicles that leave the region due to fleet rebalancing, while $\sum_{j\in\mathcal{V}} a_{ji}x_k^j$ models rebalancing vehicles arriving at i. The quantity $-\sum_{j\in\mathcal{V}} d_k^{ij}$ models all customer-occupied vehicles departing i at time k, while $\sum_{\tau=0}^{k} \sigma_\tau^{ji,k-\tau} d_\tau^{ji}$ accounts for occupied vehicles arriving to i at time k. Finally, the term g_k^i can be used to account for unmodeled disturbances affecting the dynamics, including inaccuracies in the rebalancing coefficients a_{ij} and vehicles leaving or entering the system (e.g., drivers that stop or start driving). In our simulations, we assume that the true coefficients a_{ij} are known up to a 20% parameter estimation error, and thus we model the known coefficients by perturbing the true coefficients with uniform error whose distribution has support within $\pm 20\%$ from the nominal values. Further, we assume that the travel times between regions (i.e., the scalars $\sigma_k^{ij,\tau}$) are unknown or difficult to estimate, and we incorporate all unknown terms in the exogenous signal w_k^i. Because (44) describes a mass conservation law, the dynamics (44) define a compartmental model that is marginally stable [50]. For this reason, we define the state differences $\tilde{x}_k^i := x_k^i - x_k^{i+1}$ for all $k \in \mathbb{Z}_{>0}, i \in \mathcal{V}$. In these new variables, (44) defines a $(n-1)$-dimensional system that is asymptotically stable and thus satisfies Assumption 1.

We formulate the RSP's objectives of selecting the price of rides in order to maximize its profit as the following optimization problem to be solved at every k:

$$\begin{aligned} \max_{p,x,d} \quad & \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}} p^{ij}d^{ij} - c^{ij}d^{ij} - \varrho\|x\|^2, \\ \text{s.t.} \quad & 0 = -\sum_{j\in\mathcal{V}} a_{ij}x^i + \sum_{j\in\mathcal{V}} a_{ji}x^j - \sum_{j\in\mathcal{V}} d^{ij} + w_k^i, \\ & d^{ij} = \delta_k^{ij}\left(1 - \theta^{ij}p^{ij}/p_{\max}^{ij}\right), \\ & d^{ij} \geq 0,\ x^i \geq 0, \quad \forall i, j, \in \mathcal{V}, \end{aligned} \tag{45}$$

where p, x, d denote the vectors obtained by stacking p^{ij}, x^i, and d^{ij}, for all $i, j \in \mathcal{V}$, respectively, $p^{ij}d^{ij}$ models the RSP

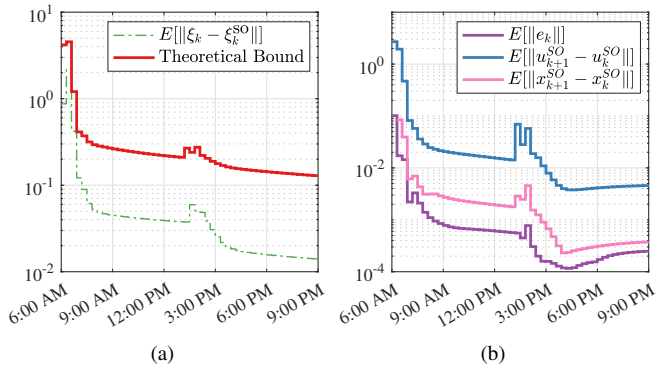


Fig. 5. (a) Numerical error and error bound from Theorem 7.1 and (b) terms characterizing the error bound. Curves illustrate the average over 100 realizations of the noise terms. Fluctuation at around 1:00PM is caused by a drop in ride demands (see Fig. 6 (top panel)).

earnings from serving the demand d^{ij}, the quantity $c^{ij}d^{ij}$, $c^{ij} \in \mathbb{R}_{>0}$, models the cost of routing vehicles from i to j. Finally, $\varrho\|x\|$, $\varrho \in \mathbb{R}_{>0}$, describes the RSP's objective of maximizing fleet utilization. We notice that this term robustifies the control objective by guaranteeing that the solution of (45) is optimal for any instantaneous value of the noise terms.

To solve the optimization problems we employ the projected controller (23) to account for constraints, with controller gain $\eta = 0.9/\mu$, where $\mu = \min_{i,j,k} \delta_k^{ij}\theta^{ij}/p_{\max}^{ij}$ is the strong convexity constant of the cost in (45). All experiments were performed using `Matlab 2019a`, ride demands and locations were estimated by using the Taxi and Limousine Commission (TLC) data from New York City [51] for March 1, 2019 between 6:00 AM and 9:00 PM. Note that the available demand data does not describe the potential rides, but rather the realized ones. Although this data may not reflect the true demand, it is often used as an approximation (see, e.g., [49]).

Fig. 5(a) illustrates the numerical tracking error and the error bound characterized in Theorem 7.1, and Fig. 5(b) presents a breakdown of the terms characterizing the tracking error. The plots show that during the initial transient the tracking error quickly decreases, up to a steady-state value of the order 10^{-2}, thus validating the conclusions drawn in Theorem 7.1. Fig. 5(b) showcases that the tracking error does not further decrease beyond such steady-state error because the optimizer (u_k^{so}, x_k^{so}) is changing over time. The sudden increase in error that occurs at around 1:00 PM can be interpreted by means of Fig 3(b), which shows that the price of rides, at this time of the day, must decrease consistently since the network experiences a drop in ride demands (see Fig. 6, top panel).

In Fig. 6, we compare the performance of the online optimization method with a fixed-pricing policy, whereby the RSP selects a fixed price for all rides, corresponding to a 25% profit from the operational cost of the fleet. A 25% profit was selected as the maximum profit that allows the RSP to serve the peak of demand with the available fleet. Fig. 6, second panel from the top, shows that by using the adaptive pricing policy the RSP always accept a higher number of rides; Fig. 6, third panel, shows that the RSP profit is always higher under the adaptive pricing policy except at the peak of demand, which can be interpreted as an optimistic situation

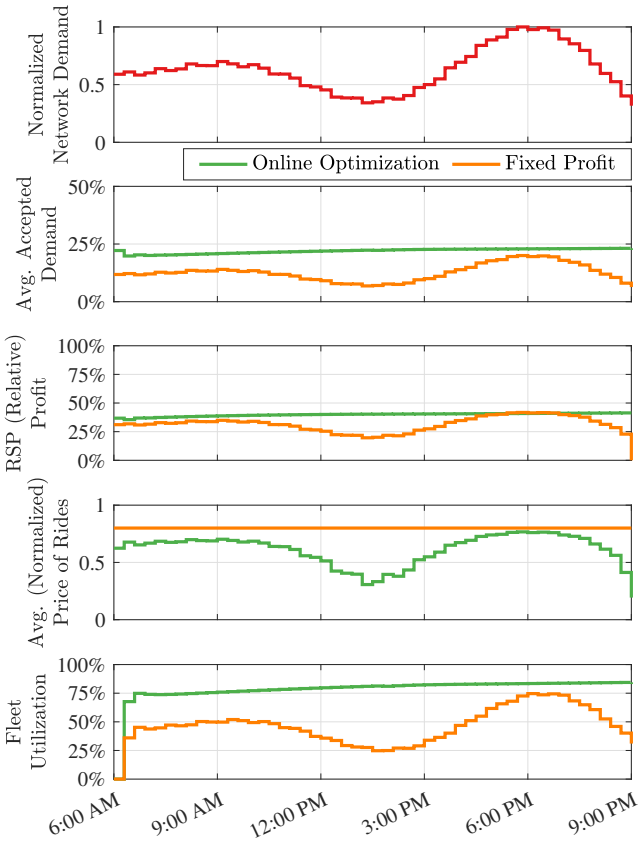


Fig. 6. Comparison between optimized pricing policy and fixed-pricing policy. In the fixed-pricing policy, the RSP sets prices to guarantee a 25% profit from the operational cost of the fleet. All lines show trajectories averaged over 100 realizations of the simulation and over the regions of the network (Fig. 4). Network demand data was derived from [51] for March 1, 2019, and normalized by its maximum value for illustration purposes. Fleet size was set to 25% of the maximum demand. The RSP profit (third panel) was normalized by the maximum profit achievable if the fleet had an infinite number of vehicles. The average price of rides (fourth panel) was normalized by the maximum willingness to pay $p^{\max}$. As shown in the middle panel (RSP relative profit), our methods outperform fixed profit policies at all times of the day, and the maximum benefit is obtained outside of peak-hours.

where suboptimal pricing still leads to high utilization of the fleet; Fig. 6, fourth panel shows that the adaptive policy adjusts the price of rides based on the instantaneous demand, and showcases that our optimal pricing policy tends to reduce the price of rides in the interest of maximizing fleet utilization; finally, Fig. 6, bottom panel shows that the optimized pricing policy always results in a higher fleet utilization.

IX. Conclusions

We have proposed a gradient-like controller to regulate a stochastic dynamical system to the solution trajectory of a stochastic and time-varying optimization problem. We showed how the controller can be synthesized by only having access to the gradient of the cost function of the optimization problem and possibly noisy input-output data generated by the open-loop system, thus overcoming the need for system identification. Overall, our work demonstrates for the first time that online optimization controllers can be synthesized directly from data, and that online stochastic optimization techniques can be used to control dynamic systems. This opens up several

exciting opportunities for future work, including extensions to scenarios where the control method guarantees persistence of excitation, to problems of simultaneous stabilization and steady-state optimization, the generalization to scenarios with distributed sensing, computation, and communication, as well as accounting for cases where costs are not strongly convex.

REFERENCES

[1] G. Bianchin, M. Vaquero, J. Cortés, and E. Dall'Anese, "Data-driven synthesis of optimization-based controllers for regulation of unknown linear systems," in *IEEE Conf. on Decision and Control*, Austin, TX, Dec. 2021, pp. 5783–5788.

[2] S. Menta, A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Stability of dynamic feedback optimization with applications to power systems," in *Annual Conf. on Communication, Control, and Computing*, Oct. 2018, pp. 136–143.

[3] M. Colombino, E. Dall'Anese, and A. Bernstein, "Online optimization as a feedback controller: Stability and tracking," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 1, pp. 422–432, 2020.

[4] G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese, "Time-varying optimization of LTI systems via projected primal-dual gradient flows," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 1, pp. 474–486, Mar. 2022.

[5] M. Nonhoff and M. A. Müller, "Online gradient descent for linear dynamical systems," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 945–952, 2020.

[6] G. Bianchin, E. Dall'Anese, J. I. Poveda, D. Jacobson, E. J. Carlton, and A. Buchwald, "Novel use of online optimization in a mathematical model of COVID-19 to guide the relaxation of pandemic mitigation measures," *Scientific Reports*, vol. 4731, no. 12, Mar. 2022.

[7] A. Jokic, M. Lazar, and P. van den Bosch, "On constrained steady-state regulation: Dynamic KKT controllers," *IEEE Transactions on Automatic Control*, vol. 54, no. 9, pp. 2250–2254, 2009.

[8] F. Brunner, H.-B. Dürr, and C. Ebenbauer, "Feedback design for multi-agent systems: A saddle point approach," in *IEEE Conf. on Decision and Control*, 2012, pp. 3783–3789.

[9] G. Bianchin, J. I. Poveda, and E. Dall'Anese, "Online optimization of switched LTI systems using continuous-time and hybrid accelerated gradient flows," *Automatica*, vol. 146, p. 110579, Aug. 2022.

[10] G. Belgioioso, D. Liao-McPherson, M. H. de Badyn, S. Bolognani, J. Lygeros, and F. Dörfler, "Sampled-data online feedback equilibrium seeking: Stability and tracking," in *IEEE Conf. on Decision and Control*, 2021, pp. 2702–2708.

[11] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Timescale separation in autonomous optimization," *IEEE Transactions on Automatic Control*, vol. 66, no. 2, pp. 611–624, 2021.

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

[13] T. Maupong and P. Rapisarda, "Data-driven control: A behavioral approach," *Systems & Control Letters*, vol. 101, pp. 37–43, 2017.

[14] C. De Persis and P. Tesi, "Formulas for data-driven control: Stabilization, optimality, and robustness," *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2019.

[15] S. Talebi, S. Alemzadeh, N. Rahimi, and M. Mesbahi, "Online regulation of unstable LTI systems from a single trajectory," *arXiv preprint*, 2020, arXiv:2006.00125.

[16] J. Coulson, J. Lygeros, and F. Dörfler, "Data-enabled predictive control: In the shallows of the DeePC," in *European Control Conference*, 2019, pp. 307–312.

[17] J. Berberich, J. Koehler, M. A. Müller, and F. Allgöwer, "Data-driven model predictive control with stability and robustness guarantees," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1702–1717, 2021.

[18] G. Baggio, V. Katewa, and F. Pasqualetti, "Data-driven minimum-energy controls for linear systems," *IEEE Control Systems Letters*, vol. 3, no. 3, pp. 589–594, 2019.

[19] L. Xu, M. Turan Sahin, B. Guo, and G. Ferrari-Trecate, "A data-driven convex programming approach to worst-case robust tracking controller design," *arXiv preprint*, 2021, arXiv:2102.11918.

[20] A. Allibhoy and J. Cortés, "Data-based receding horizon control of linear network systems," *IEEE Control Systems Letters*, vol. 5, no. 4, pp. 1207–1212, 2020.

[21] J. Berberich and F. Allgöwer, "A trajectory-based framework for data-driven system analysis and control," in *European Control Conference*, 2020, pp. 1365–1370.

[22] M. Guo, C. De Persis, and P. Tesi, "Data-driven stabilization of nonlinear polynomial systems with noisy data," *IEEE Transactions on Automatic Control*, 2022, (early access) arXiv:2011.07833.

[23] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016.

[24] S. Bolognani and S. Zampieri, "A distributed control strategy for reactive power compensation in smart microgrids," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2818–2833, 2013.

[25] A. Bernstein, E. Dall'Anese, and A. Simonetto, "Online primal-dual methods with measurement feedback for time-varying convex optimization," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 1978–1991, 2019.

[26] C.-Y. Chang, M. Colombino, J. Cortés, and E. Dall'Anese, "Saddle-flow dynamics for distributed feedback-based optimization," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 948–953, 2019.

[27] L. S. P. Lawrence, Z. E. Nelson, E. Mallada, and J. W. Simpson-Porco, "Optimal steady-state control for linear time-invariant systems," in *IEEE Conf. on Decision and Control*, Dec. 2018, pp. 3251–3257.

[28] D. Li, D. Fooladivanda, and S. Martínez, "Online optimization and learning in uncertain dynamical environments with performance guarantees," *arXiv preprint*, 2021, arXiv:2102.09111.

[29] K. Hirata, J. Hespanha, and K. Uchida, "Real-time pricing leading to optimal operation under distributed decision makings," in *American Control Conference*, 2014, pp. 1925–1932.

[30] M. Nonhoff and M. A. Müller, "Data-driven online convex optimization for control of dynamical systems," in *IEEE Conf. on Decision and Control*, 2021, pp. 3640–3645.

[31] J. C. Perdomo, T. Zrnic, C. Mendler-Dünner, and M. Hardt, "Performative prediction," in *International Conference on Machine Learning*, 2020, pp. 7599–7609.

[32] C. Mendler-Dünner, J. C. Perdomo, T. Zrnic, and M. Hardt, "Stochastic optimization for performative prediction," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4929–4939, 2020.

[33] D. Drusvyatskiy and L. Xiao, "Stochastic optimization with decision-dependent distributions," *Mathematics of Operations Research*, 2022.

[34] M. Yin, A. Iannelli, and R. S. Smith, "Maximum likelihood estimation in data-driven modeling and control," *IEEE Transactions on Automatic Control*, vol. 68, no. 1, pp. 317–328, 2021.

[35] H. J. van Waarde, M. K. Camlibel, and M. Mesbahi, "From noisy data to feedback controllers: non-conservative design via a matrix S-lemma," *IEEE Transactions on Automatic Control*, vol. 67, no. 1, pp. 162–175, 2022.

[36] A. Bisoffi, C. De Persis, and P. Tesi, "Trade-offs in learning controllers from noisy data," *Systems & Control Letters*, vol. 154, p. 104985, 2021.

[37] S. Bubeck, "Convex optimization: Algorithms and complexity," *Foundations and Trends Machine Learning*, vol. 8, no. 3–4, p. 231–357, Nov. 2015.

[38] L. V. Kantorovich and S. G. Rubinshtein, "On a space of totally additive functions," *Vestnik Leningradskogo Universiteta*, vol. 13, no. 7, pp. 52–59, 1958.

[39] E. Davison, "The robust control of a servomechanism problem for linear time-invariant multivariable systems," *IEEE Transactions on Automatic Control*, vol. 21, no. 1, pp. 25–34, 1976.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[41] H. Karimi, J. Nutini, and M. Schmidt, "Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2016, pp. 795–811.

[42] I. Markovsky and P. Rapisarda, "Data-driven simulation and control," *International Journal of Control*, vol. 81, no. 12, pp. 1946–1959, 2008.

[43] F. Dorfler, J. Coulson, and I. Markovsky, "Bridging direct & indirect data-driven control formulations via regularizations and relaxations," *IEEE Transactions on Automatic Control*, 2022, (Early access).

[44] V. Krishnan and F. Pasqualetti, "On direct vs indirect data-driven predictive control," in *IEEE Conf. on Decision and Control*, 2021, pp. 736–741.

[45] L. Ljung, *System Identification: Theory for the User*. Prentice Hall, 1999.

[46] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, "Online optimization: Competing with dynamic comparators," in *Artificial Intelligence and Statistics*, 2015, pp. 398–406.

[47] E. Dall'Anese, A. Simonetto, S. Becker, and L. Madden, "Optimization and learning with information streams: Time-varying algorithms and applications," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 71–83, 2020.
[48] A. Khaled and P. Richtárik, "Better theory for SGD in the nonconvex world," *arXiv preprint*, 2020, arXiv:2002.03329.
[49] B. Turan and M. Alizadeh, "Competition in electric autonomous mobility on demand systems," *IEEE Transactions on Control of Network Systems*, 2021, in press.
[50] W. M. Haddad, V. S. Chellaboina, and E. August, "Stability and dissipativity theory for discrete-time non-negative and compartmental dynamical systems," *International Journal of Control*, vol. 76, no. 18, pp. 1845–1861, 2003.
[51] "TLC trip record data," https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page, [Online; accessed 12-Aug-2021].

Gianluca Bianchin (S'15, M'20) is an Assistant Professor at the ICTEAM Institute and Department of Mathematical Engineering at the Université Catholique de Louvain, Belgium. He received the Doctor of Philosophy degree in Mechanical Engineering from the University of California Riverside in 2020, the Laurea degree in Information Engineering, and the Laurea Magistrale degree (Summa Cum Laude) in Controls Engineering from the University of Padova, Italy, in 2012 and 2014, respectively. He was a Postdoctoral Scholar in the ECEE Department at the University of Colorado Boulder from 2020 to 2022. Prof. Bianchin is the recipient of the Dissertation Year Award and the Dean's Distinguished Fellowship Award from the University of California Riverside; his work on secure navigation was the Elsevier Journal of Automatica Editor's choice of the month in February 2020. His research interests include dynamical systems and control theory and their applications in traffic control and complex network control.

Miguel Vaquero was born in Galicia, Spain. He received his Licenciatura and Master's degree in mathematics from the Universidad de Santiago de Compostela, Spain and the Ph.D. degree in mathematics from Instituto de Ciencias Matemáticas (ICMAT), Spain in 2015. He was then a postdoctoral scholar working on the ERC project "Invariant Submanifolds in Dynamical Systems and PDE" also at ICMAT. From 2017 to 2020, he was a postdoctoral scholar in the Department of Mechanical and Aerospace Engineering of UC San Diego. Since January 2021, he has been an Assistant Professor in the School of Human Sciences and Technology at IE University, Madrid, Spain. His interests include optimization, dynamical systems, control theory, machine learning, and geometric mechanics.

Jorge Cortés (M'02, SM'06, F'14) received the Licenciatura degree in mathematics from Universidad de Zaragoza, Zaragoza, Spain, in 1997, and the Ph.D. degree in engineering mathematics from Universidad Carlos III de Madrid, Madrid, Spain, in 2001. He held postdoctoral positions with the University of Twente, Twente, The Netherlands, and the University of Illinois at Urbana-Champaign, Urbana, IL, USA. He was an Assistant Professor with the Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, USA, from 2004 to 2007. He is currently a Professor in the Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA, USA. He is a Fellow of IEEE and SIAM. He is the author of Geometric, Control and Numerical Aspects of Nonholonomic Systems (Springer-Verlag, 2002) and co-author (together with F. Bullo and S. Martínez) of Distributed Control of Robotic Networks (Princeton University Press, 2009). At the IEEE Control Systems Society, he has been a Distinguished Lecturer (2010-2014) and an elected member (2018-2020) of its Board of Governors, and is currently its Director of Operations.

Emiliano Dall'Anese is an Assistant Professor in the Department of Electrical, Computer, and Energy Engineering at the University of Colorado Boulder, and an affiliate Faculty with the Department of Applied Mathematics. He received the Ph.D. in Information Engineering from the Department of Information Engineering, University of Padova, Italy, in 2011. From January 2011 to November 2014, he was a Postdoctoral Associate at the Department of Electrical and Computer Engineering of the University of Minnesota, and from December 2014 to July 2018 he was a Senior Researcher at the National Renewable Energy Laboratory. He received the National Science Foundation CAREER Award in 2020, and the IEEE PES Prize Paper Award in 2021.