Automatica 158 (2023) 111265

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Robust and scalable distributed recursive least squares[☆]

Ilario Antonio Azzollini[a], Michelangelo Bin[a,b], Lorenzo Marconi[a], Thomas Parisini[b,c,d,*]

[a] Department of Electrical, Electronic, and Information Engineering, University of Bologna, Bologna, Italy
[b] Department of Electrical and Electronic Engineering, Imperial College London, London, UK
[c] KIOS Research and Innovation Center of Excellence, University of Cyprus, Cyprus
[d] Department of Engineering and Architecture, University of Trieste, Trieste, Italy

ARTICLE INFO

Article history:
Received 11 June 2022
Received in revised form 11 March 2023
Accepted 21 July 2023
Available online 13 September 2023

Keywords:
Distributed least squares
Robust estimation
Distributed optimization

ABSTRACT

We consider a problem of robust estimation over a network in an errors-in-variables context. Each agent measures noisy samples of a local pair of signals related by a linear regression defined by a common unknown parameter, and the agents must cooperate to find the unknown parameter in presence of uncertainty affecting both the regressor and the regressand variables. We propose a recursive least squares estimation method providing global exponential convergence to the unknown parameter in absence of uncertainty, and robust stability of the estimate, formalized in terms of input-to-state stability, in presence of uncertainty affecting all the variables. The result relies on a cooperative excitation assumption that is proved to be strictly weaker than persistency of excitation of each local data set. The proposed estimator is validated on an adaptive road pricing application.

1. Introduction

1.1. Problem overview and contribution

We consider a set of $n \in \mathbb{N}$ agents. Each agent i measures noisy samples of a pair (y_i, ϕ_i) of signals related by a linear regression of the form

$$y_i(t) = \phi_i(t)^\top \theta, \tag{1}$$

in which θ is a *common* unknown parameter. Agents exchange information through a communication network, possibly disconnected and asymmetric. In this setting, we consider the problem of distributed online asymptotic estimation of the unknown parameter θ. While this problem can be in principle solved with n independent local algorithms, each one trying to estimate θ from the local samples, communication permits agents to benefit from the information gathered by the other agents, and thus ensures faster convergence under excitation conditions that are *strictly weaker* than persistency of excitation of each local data set. Hence, looking for a distributed design is well motivated in all those contexts where measurements are difficult or expensive, when sensors are spatially distributed by construction, or simply when faster convergence is needed. For ease of exposition, we focus on the "single-variable" case where $\theta \in \mathbb{R}^{n_\theta}$ for some $n_\theta \in \mathbb{N}$ and, for each $i = 1, \dots, n$, $y_i : \mathbb{N} \to \mathbb{R}$ and $\phi_i : \mathbb{N} \to \mathbb{R}^{n_\theta}$. Nevertheless, we remark that the proposed approach easily extends to a multi-variable setting, where $y_i(t) \in \mathbb{R}^m$, $\phi_i(t) \in \mathbb{R}^{n_\theta \times m}$ and $\theta \in \mathbb{R}^{n_\theta \times m}$ for some $m > 1$, by suitably concatenating m single-variable solutions. We consider a discrete-time setup where, at each step, every agent collects its new samples, exchanges its local state with a subset of other agents, and then updates its local estimate of θ. This is repeated over an infinite time horizon. We suppose that the measured samples are corrupted by additive disturbances on which we make no prior assumption. These disturbances can model measurement noise, unmodeled dynamics/terms in (1), or any other sources of uncertainty in the measurement process. As disturbances affect both y_i and ϕ_i, we are in an *errors-in-variables* context (Söderström, 2007). We propose a decentralized estimation law by which each agent can asymptotically estimate the common unknown parameter θ in a robust way. In particular, we prove the existence and robust stability of an aggregate steady-state trajectory that, if a "cooperative excitation" condition holds, is associated with a correct estimation of θ. Robustness is meant in the sense of *input-to-output* and *input-to-state stability* (ISS) (Sontag & Wang, 1996) with respect to the exogenous disturbances, and the main result

☆ M.B. and T.P. acknowledge support from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement no. 739551 (KIOS CoE). T.P also acknowledges support from the Italian Ministry for Research in the framework of the 2017 Program for Research Projects of National Interest (PRIN), Grant no. 2017YKXYXJ. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wei Xing Zheng under the direction of Editor Alessandro Chiuso.

* Corresponding author.
E-mail addresses: ilario.azzollini92@gmail.com (I.A. Azzollini), michelangelo.bin@unibo.it (M. Bin), lorenzo.marconi@unibo.it (L. Marconi), t.parisini@imperial.ac.uk (T. Parisini).

https://doi.org/10.1016/j.automatica.2023.111265

of the article states that each agent's asymptotic estimation error of θ is bounded by a continuous function of the asymptotic magnitude of the disturbances. In particular, in case the disturbances vanish, exact convergence is achieved.

1.2. Related works

Distributed estimation is a well-developed research field boasting a large number of contributions from the control and signal processing community (see, e.g., Sayed, Tu, Chen, Zhao, & Towfic, 2013 for an overview). In particular, regarding the problem described in Section 1.1, the vast majority of the approaches present in the literature are based on distributed versions of the *least squares* (Breschi, Bemporad, & Kolmanovsky, 2020; Mateos & Giannakis, 2012; Mateos, Schizas, & Giannakis, 2009; Xie, Zhang, & Guo, 2021; Yu, Zhao, de Lamare, Zakharov, & Lu, 2019), the *least mean squares* (Lopes & Sayed, 2008; Schizas, Mateos, & Giannakis, 2009; Xie & Guo, 2018a, 2018b, 2018c), or the *total least squares* (Li, Zhao, & Lv, 2021) methods. Moreover, in addition to the specific criterion to be optimized, the structure of the existing algorithms also depends on the network's constraints, these being posed in terms of connection topology and on the type and quality of exchangeable information. In particular, one can divide the existing approaches into three types: *consensus* (Breschi et al., 2020; Johansson, Keviczky, Johansson, & Johansson, 2008; Mateos & Giannakis, 2012; Mateos et al., 2009; Schizas et al., 2009; Xie & Guo, 2018b, 2018c), *diffusion* (Cattivelli, Lopes, & Sayed, 2008; Chen & Sayed, 2012; Lopes & Sayed, 2008; Sayed et al., 2013; Xie & Guo, 2018a; Xie et al., 2021; Yu et al., 2019), and *incremental* (Cattivelli & Sayed, 2010; Lopes & Sayed, 2007) methods. Consensus methods are based on a consensus mechanism aimed at synchronizing the agents' estimates of the parameter θ. Diffusion methods, instead, are based on an *aggregation* step aimed to process the information collected from the neighbors and merge it with the local measurements. Each agent then computes its own local estimate of the unknown parameter θ privately, without any consensus protocol. At each iteration, the aggregation step may either be performed before the computation of the estimate of θ ("combine then adapt") or after ("adapt then combine"). Generally speaking, consensus and diffusion methods have no structural differences in terms of computational complexity. However, diffusion strategies are typically preferable over consensus strategies since they converge faster and reach lower mean-square error deviations (Sayed et al., 2013; Tu & Sayed, 2012). Moreover, diffusion methods do not require the exchange of the local estimates of θ and are therefore preferable for privacy preservation. In addition, diffusion methods generally lead to enhanced stability properties (Sayed et al., 2013). Both consensus and diffusion methods, however, may require the exchange of a significant amount of data, which have to be processed locally in real time. Incremental methods aim to mitigate this problem for certain communication structures by exploiting cyclic paths (Cattivelli & Sayed, 2010; Lopes & Sayed, 2007). However, determining a cyclic path is NP-hard, and such paths are not robust with respect to link and node failures, or to changes in the communication topology (Sayed et al., 2013). Other approaches related to this work include the "partial diffusion" approaches of Arablouei, Doğançay, Werner, and Huang (2014) and Rastegarnia (2019), which aim to mitigate the exchange of data in diffusion methods, the sparse recursive least square method of Liu, Liu, and Li (2014), aimed to reduce computational complexity, and the data-adaptive censoring method proposed in Wang, Yu, Ling, Berberidis, and Giannakis (2018), which could be beneficial in large-scale networks. For what concerns the uncertainty in the measurement of the signals (y_i, ϕ_i) in (1), most of the existing approaches model it in terms of stochastic additive disturbances, and rely on the typical assumptions of independence, stationarity, and Gaussianity. Moreover, except for rare exceptions such as (Li et al., 2021), disturbances are only assumed to affect y_i and not ϕ_i, unlike this article. Notably, the recent distributed least mean square methods of Xie and Guo (2018a, 2018b, 2018c) and the distributed least squares method of Xie et al. (2021) relax the aforementioned assumptions, and thus are potentially applicable to more involved stochastic feedback systems. In addition, they introduce the concept of *cooperative excitation* as a distributed relaxation of classical persistency of excitation. A notion of cooperative persistency of excitation also appeared before in the broader context of distributed adaptive control, see, e.g., Chen, Hua and Sam Ge (2014), Chen, Wen, Hua and Sun (2014) and Javed, Poveda, and Chen (2022). An equivalent cooperative excitation property, which depends on the data acquired by all the agents in the network, is also used in this article to guarantee the main stability and robustness results.

In terms of the previously-defined terminology, the algorithm proposed in this paper can be categorized as a combine-then-adapt diffusion method. As anticipated in Section 1.1, contrary to most existing approaches, we consider that both y_i and ϕ_i are affected by generic disturbance terms on which we make no statistical assumption, and we shift the focus from unbiasedness and consistency to stability and robustness (in terms of ISS). This control-oriented methodology enables the use of distributed least squares methods in complex feedback control systems, since ISS opens the door to canonical nonlinear control techniques such as small-gain methods (Jiang, Teel, & Praly, 1994). In a centralized setting, recursive least squares algorithms with ISS guarantees are studied in Bin (2022). This article parallels such results in a distributed setting.

1.3. Notation

We denote by $\mathbb{R}$ and $\mathbb{N}$ the set of real and natural numbers, respectively ($0 \in \mathbb{N}$). By $\subset$ we denote non-strict inclusion. Given a family $(X_i)_{i\in I}$ of sets X_i, the elements of the Cartesian product $\prod_{i\in I} X_i$ are denoted by $(x_i)_{i\in I}$. These shall be interpreted as column/row arrangements as needed. If A is a set, A^n denotes the n-fold Cartesian product of A. If $(S, \geq)$ is a preordered set, for every $s \in S$ we let $S_{\geq s} := \{z \in S : z \geq s\}$. If $A_1, \ldots, A_n$ are operators, we denote by $\operatorname{diag}(A_1, \ldots, A_n)$ their diagonal concatenation. By $\otimes$ we denote the Kronecker product and by $\sigma(A)$ the spectrum of an operator A. For $p = 1, \ldots, \infty$, $|\cdot|_p$ denotes the vector, or operator induced, p-norm. We denote by $\mathbb{P}_n$ the set of positive semi-definite symmetric operators on $\mathbb{R}^n$. We write $A \geq B$ if $A - B \in \mathbb{P}_n$. The n-dimensional identity operator is denoted by I_n and the $n \times m$ zero operator by $0_{n\times m}$. Dimensions are omitted when clear. A function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class-$\mathcal{K}$ ($\gamma \in \mathcal{K}$) if it is continuous, strictly increasing, and $\gamma(0) = 0$. We denote by $\cdot^+$ the shift operator, i.e. $x^+(t) = x(t+1)$.

2. The framework

2.1. Samples acquisition

Let $\mathcal{N} := \{1, \ldots, n\}$ denote the set of agents. At each time $t \in \mathbb{N}$, each agent $i \in \mathcal{N}$ samples the signals (y_i, ϕ_i) of the regression (1) and obtains the samples

$$\xi_i(t) = y_i(t) + \delta_{y,i}(t), \quad \varphi_i(t) = \phi_i(t) + \delta_{\phi,i}(t), \tag{2}$$

in which $\delta_i := (\delta_{y,i}, \delta_{\phi,i}) : \mathbb{N} \to \mathbb{R}^{n_\delta}$, $n_\delta := n_\theta + 1$, represents the exogenous perturbations adding to the measurements of y_i and ϕ_i and modeling, for example, noise affecting the measurement process and/or unmodeled dynamics. Due to the presence of δ_i affecting both the regressor ϕ_i and the regressand y_i, estimating θ is an *errors-in-variables* problem (Söderström, 2007). In the remainder of the paper, we let $y := (y_i)_{i\in\mathcal{N}}$, $\phi := (\phi_i)_{i\in\mathcal{N}}$, $\delta := (\delta_i)_{i\in\mathcal{N}}$, $\xi := (\xi_i)_{i\in\mathcal{N}}$, $\varphi := (\varphi_i)_{i\in\mathcal{N}}$, and we make the following uniform boundedness assumption.

Assumption 1. There exist $\bar{y}, \bar{\phi} > 0$, such that $|y_i(t)| \le \bar{y}$ and $|\phi_i(t)|_\infty \le \bar{\phi}$ for every $i \in \mathcal{N}$ and $t \in \mathbb{N}$.

2.2. Information network

We assume that agents can exchange information over an information network formally described by a family $\mathcal{C} = \{I_i\}_{i\in\mathcal{N}}$ of sets $I_i \subset \mathcal{N}$ satisfying $i \in I_i$. We call $\mathcal{C}$ the *information network*, and the set I_i the *(inward) neighborhood* of i. Each agent i can receive information by all the agents $k \in I_i$, and can send information to all the agents in the set $O_i := \{k \in \mathcal{N} : i \in I_k\}$. It is not required that $k \in I_i$ implies $i \in I_k$, so the network can be directed. We denote by d_i the cardinality of I_i, and we associate with $\mathcal{C}$ the *adjacency matrix* $A \in \mathbb{R}^{n\times n}$, defined by letting $A_{ik} = 1$ if $k \in I_i$ and $A_{ik} = 0$ otherwise, where A_{ik} denotes the (i, k)th entry of A. We stress that $A_{ii} = 1$ for all $i \in \mathcal{N}$. Finally, we introduce the matrix $\Lambda := \operatorname{diag}\big(d_1, \dots, d_n\big)^{-1} A$.

2.3. Problem statement

We consider the problem of designing, for each agent $i \in \mathcal{N}$, a recursive procedure that exploits the samples acquired by (2) and the information coming from the neighboring agents to determine, asymptotically, a "well-behaved" estimate $\hat{\theta}_i$ of the unknown common parameter θ in (1), in the sense that the following properties hold: **(i)** *Nominal Exactness:* In absence of disturbances, i.e. when $\delta = 0$, $\lim_{t\to\infty} \hat{\theta}_i(t) = \theta$ for all $i \in \mathcal{N}$. **(ii)** *Robustness:* If $\delta \neq 0$, the asymptotic estimation errors must be related, in a continuous manner, to the "asymptotic magnitude" of δ. Formally, there must exist $\kappa \in \mathcal{K}$ such that $\limsup_{t\to\infty} |\hat{\theta}_i(t) - \theta| \le \kappa(\limsup_{t\to\infty} |\delta(t)|)$ for all $i \in \mathcal{N}$ and all bounded δ. **(iii)** *Decentralization:* The update law of $\hat{\theta}_i$ must only depend on the current samples produced by (2) and the information communicated by the agents $j \in I_i$. **(iv)** *Scalability:* The update laws must be independent from "centralized" parameters or quantities.

3. Distributed recursive least squares

3.1. The update laws

To approach the problem illustrated in Section 2.3, we propose the following update law for each agent $i \in \mathcal{N}$:

$$\begin{aligned} \Psi_i^+ &= \frac{\mu_i}{d_i} \sum_{k\in I_i} \Psi_k + (1-\mu_i)\varphi_i\varphi_i^\top, \\ \eta_i^+ &= \frac{\mu_i}{d_i} \sum_{k\in I_i} \eta_k + (1-\mu_i)\varphi_i\xi_i, \end{aligned} \tag{3}$$

with output

$$\hat{\theta}_i = \gamma_i\big(\Psi_i, \eta_i\big). \tag{4}$$

In (3)–(4), Ψ_i and η_i are the state variables associated with agent i and take their values in $\mathbb{P}_{n_\theta}$ and $\mathbb{R}^{n_\theta}$ respectively, the output $\hat{\theta}_i$ is the estimate agent i has on θ and it takes values in $\mathbb{R}^{n_\theta}$, φ_i and ξ_i are the samples produced by (2), the parameters $\mu_i \in (0, 1)$ are arbitrarily chosen, and the functions $\gamma_i : \mathbb{P}_{n_\theta} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_\theta}$ are degrees of freedom fixed later in Section 3.5. We also let $\Psi := (\Psi_i)_{i\in\mathcal{N}}$ and $\eta := (\eta_i)_{i\in\mathcal{N}}$.

3.2. The aggregate system

Let $\Phi : (\mathbb{R}^{n_\theta})^n \to (\mathbb{P}_{n_\theta})^n$ and $\Xi : (\mathbb{R}^{n_\theta})^n \times \mathbb{R}^n \to (\mathbb{R}^{n_\theta})^n$ be functions mapping $\varphi = (\varphi_i)_{i\in\mathcal{N}} \in (\mathbb{R}^{n_\theta})^n$ and $(\varphi, \xi) = ((\varphi_i)_{i\in\mathcal{N}}, (\xi_i)_{i\in\mathcal{N}}) \in (\mathbb{R}^{n_\theta})^n \times \mathbb{R}^n$ to $\Phi(\varphi) := (\varphi_i\varphi_i^\top)_{i\in\mathcal{N}}$ and $\Xi(\varphi, \xi) := (\varphi_i\xi_i)_{i\in\mathcal{N}}$, respectively. Let $W := \operatorname{diag}(\mu_1, \dots, \mu_n)$, $F := W\Lambda \otimes I_{n_\theta}$, and $G := (I_n - W) \otimes I_{n_\theta}$. Then, the composition of (2) and (3) is a system with input (y, ϕ, δ), state variable $x := (\Psi, \eta)$ ranging in the state space $\mathcal{X} := (\mathbb{P}_{n_\theta})^n \times (\mathbb{R}^{n_\theta})^n$, output $\hat{\theta} := (\hat{\theta}_i)_{i\in\mathcal{N}}$, and its dynamics is described by

$$x^+ = g(x, y, \phi, \delta) \tag{5}$$

where

$$\begin{aligned} g(x, y, \phi, \delta) :=& (F\Psi, F\eta) \\ &+ \big(G\Phi(\phi + \delta_\phi),\ G\Xi(\phi + \delta_\phi, y + \delta_y)\big) \end{aligned} \tag{6}$$

with $\delta_\phi := (\delta_{\phi,i})_{i\in\mathcal{N}}$ and $\delta_y := (\delta_{y,i})_{i\in\mathcal{N}}$. In (6), Ψ, η, $\Phi(\cdot)$, and $\Xi(\cdot)$ are interpreted as column concatenations. We define on $\mathcal{X}$ the functional $|x| := \max\big\{|\Psi|_\infty, |\eta|_\infty\big\}$.

3.3. Existence and robust stability of a steady state

In this section, we study the stability properties and the asymptotic behavior of the aggregate system (5). In particular, we show that (5) is ISS relative to a time-varying steady-state solution $t \mapsto (\Psi^\star(t), \eta^\star(t))$ defined by the unperturbed signals (ϕ, y) and with respect to the disturbance δ. Later in Section 3.4, we shall prove that, if a suitable excitation condition is fulfilled, such steady-state solution $(\Psi^\star, \eta^\star)$ is associated with a correct estimate of θ for all the agents. Let $\bar{\mu} := \max_{i=1,\dots,n} \mu_i$. Since $\mu_i \in (0, 1)$ for all $i \in \mathcal{N}$ by definition, then $\bar{\mu} < 1$. As a consequence, the operator F in (6) is Schur stable as established by the following proposition.

Proposition 2. *$|F|_\infty \le \bar{\mu}$ and, for all $\lambda \in \sigma(F)$, $|\lambda| < 1$.*

Proof. By definition of A in Section 2.2, $\sum_{k=1}^n A_{ik} = d_i$ for all $i = 1, \dots, n$. Thus, $|F|_\infty = |W\Lambda|_\infty = \max_{i=1,\dots,n} \sum_{k=1}^n \mu_i d_i^{-1} A_{ik} = \bar{\mu}$. As $|\cdot|_\infty$ is sub-multiplicative, then $\max_{\lambda\in\sigma(F)} |\lambda| \le |F|_\infty \le \bar{\mu} < 1$.

With the unperturbed signals ϕ and y we associate a signal $x^\star = (\Psi^\star, \eta^\star) : \mathbb{N} \to \mathcal{X}$ defined as

$$\begin{aligned} \Psi^\star(t) &:= \sum_{s=0}^{t-1} F^{t-s-1} G\Phi(\phi(s)), \\ \eta^\star(t) &:= \sum_{s=0}^{t-1} F^{t-s-1} G\Xi(\phi(s), y(s)). \end{aligned} \tag{7}$$

Let $(\Psi_i^\star)_{i\in\mathcal{N}}$ and $(\eta_i^\star)_{i\in\mathcal{N}}$ be such that $\Psi^\star$ and $\eta^\star$ satisfy $\Psi^\star(t) = (\Psi_i^\star(t))_{i\in\mathcal{N}}$ and $\eta^\star(t) = (\eta_i^\star(t))_{i\in\mathcal{N}}$ at each $t \in \mathbb{N}$. Then, $\Psi_i^\star(t) \in \mathbb{P}_{n_\theta}$ for all $t \in \mathbb{N}$, since $0 \in \mathbb{P}_{n_\theta}$ and $\phi_i(s)\phi_i(s)^\top \in \mathbb{P}_{n_\theta}$ for all $i \in \mathcal{N}$ and $s \in \mathbb{N}$. Moreover, we underline that $x^\star$ is only defined by the unperturbed samples of (1), and it does not depend on δ or on the aggregate state variable x. Finally, we observe that $(x^\star, (y, \phi, 0))$ is a solution pair of (5) satisfying $x^\star(0) = 0$. The forthcoming proposition establishes input-to-state stability of (5) relative to $x^\star$ and with respect to the input δ. If $\delta = 0$, the result implies asymptotic convergence of x to $x^\star$. Thus, we refer to $x^\star$ as the *ideal steady state* of x. Let $\underline{\mu} := \min_{i\in\mathcal{N}} \mu_i$ and, with $\bar{y}$ and $\bar{\phi}$ given by Assumption 1, define

$$\omega(s) := (1 - \underline{\mu})(1 - \bar{\mu})^{-1}\big(2\max\{\bar{\phi}, \bar{y}\}s + s^2\big). \tag{8}$$

Then, $\omega \in \mathcal{K}$ and the following holds (the proof is given in Appendix A).

Proposition 3. *Suppose that Assumption 1 holds. Then, there exists $\alpha > 0$ such that, for every solution pair $(x, (y, \phi, \delta))$ of (5), the following holds for all $t \in \mathbb{N}$*

$$|x(t) - x^\star(t)| \le \alpha \bar{\mu}^t |x(0) - x^\star(0)| + \omega\left(\sup_{s\in\mathbb{N}_{\le t-1}} |\delta(s)|_\infty\right).$$

Remark 4. The function ω, defined in (8), is quadratic. If the estimation scheme (2), (3) has to be used as a component in a larger control system, it might be useful to have a function ω which is sub-linear, as this is necessary to enforce small-gain like conditions with linear feedback. We remark that this can be achieved as shown in Bin (2022) by "saturating" the terms $\varphi_i\varphi_i^\top$ and $\varphi_i\xi_i$ in (3) within a ball of radius larger than $\bar{\phi}^2$ and $\max\{\bar{y}, \bar{\phi}\}^2$, respectively.

3.4. Cooperative excitation and identifiability

Existence and robust stability of the steady-state trajectory $x^\star$ is always guaranteed if the signals y and ϕ are bounded. In this section we show that, under suitable excitation conditions, $x^\star$ is also associated with a correct estimate of θ for each agent.

Lemma 5. *For all $i \in \mathcal{N}$ and $t \in \mathbb{N}$, $\Psi_i^\star(t)\theta = \eta_i^\star(t)$.*

Lemma 5, proved in Appendix B, guarantees that the sought parameter θ satisfies $\Psi_i(t)\theta = \eta_i(t)$ for all $t \in \mathbb{N}$ in the ideal steady state $x^\star$, which is exponentially attractive when $\delta = 0$. However, it does not guarantee that θ can be uniquely determined by $\Psi_i^\star(t)$ and $\eta_i^\star(t)$, as in general $\Psi_i^\star(t)$ may be singular or ill-conditioned. Identifiability of θ requires a further condition formalized in this paper in terms of the following excitation property.

Definition 6. With $(\varepsilon_i, \tau_i) \in \mathbb{R}_{>0} \times \mathbb{N}$, agent $i \in \mathcal{N}$ is said to be (ε_i, τ_i)-*excited* if

$$\forall t \ge \tau_i, \qquad \Psi_i^\star(t) \ge \varepsilon_i I. \tag{9}$$

The following proposition directly follows from Lemma 5 and connects Definition 6 and identifiability of θ.

Proposition 7. *If agent $i \in \mathcal{N}$ is (ε_i, τ_i)-excited, then for all $t \ge \tau_i$, $\Psi_i^\star(t)$ is invertible and $\theta = \Psi_i^\star(t)^{-1}\eta_i^\star(t)$.*

In the remainder of the section, we further discuss the excitation condition of Definition 6. In view of (7), for each $t \in \mathbb{N}$, we have

$$\Psi^\star(t) = \sum_{s=0}^{t-1} F^{t-s-1} G\Big(\phi_i(s)\phi_i(s)^\top\Big)_{i\in\mathcal{N}}$$

in which $(\phi_i(s))_{s=0,\dots,t-1}$ are samples of the regressor ϕ_i of agent i, and $F = W\Lambda \otimes I$ incorporates the information network. Therefore, each matrix $\Psi_i^\star(t)$ is an "aggregate" quantity given by the combination of terms of the kind $\phi_k(s)\phi_k(s)^\top$, obtained by mixing the local endogenous samples ($k = i$) with those available to the other agents of the network ($k \in \mathcal{N} \setminus \{i\}$), all properly weighted by the factors μ_i and filtered. As a consequence, the excitation condition (9) is not only a property of the regressor measured by agent i, but it depends on all the samples of all the other agents of the network. Hence, the excitation condition of Definition 6 is a "cooperative" condition. We also stress that, for each $i \in \mathcal{N}$, $\Psi_i^\star$ only depends on the information network, the unperturbed regressors $(\phi_j)_{j\in\mathcal{N}}$, and the parameters $(\mu_j)_{j\in\mathcal{N}}$. Instead, it does not depend on the states $x = (\Psi, \eta)$ of the algorithm (5) nor on the uncertainties δ. Therefore, once the information network and the parameters $(\mu_j)_{j\in\mathcal{N}}$ are fixed, (ε_i, τ_i)-excitation is an open-loop assumption on the regressors $(\phi_i)_{i\in\mathcal{N}}$. In this connection, it is worth comparing Definition 6 with the "cooperative persistency of excitation" assumption of Chen, Hua et al. (2014) (see also Chen, Wen et al., 2014 for a continuous-time version, and Javed et al., 2022 for an extension to disconnected networks). In the notation of this article, Chen, Hua et al. (2014, Def. 4) reads as

$$\exists \alpha, T \ge 0, \quad \forall t \ge 0, \qquad \sum_{s=t}^{t+T-1} \sum_{i\in\mathcal{N}} \phi_i(s)\phi_i(s)^\top \ge \alpha I. \tag{10}$$

It is not difficult to see that, if the information network is strongly connected (meaning that there is a path between every two agents), then (10) implies that each agent $i \in \mathcal{N}$ is (ε_i, τ_i)-excited for a suitable value of ε_i, τ_i, and μ_i depending on T and α in (10). The converse is also true. Namely, if (9) holds for each $i \in \mathcal{N}$, then (10) holds for suitable values of T and α. This equivalence can be extended also to the weaker notion of Javed et al. (2022) when the information network is disconnected. To ease the technical derivations, in this article it is more convenient to directly assume (9) instead of (10).

One may wonder if the contribution of communication in achieving (9) carries any advantages with respect to the case in which agent i only uses its own samples to compute an estimate of θ. In our setting, this is a well-posed question, since the information network has no connectivity requirements to satisfy (Proposition 2 holds for every adjacency matrix A). In particular, for every $i \in \mathcal{N}$, define $\Psi_i^L : \mathbb{N} \to \mathbb{P}_{n_\theta}$ by letting $\Psi_i^L(0) := 0$ and

$$\Psi_i^L(t+1) = \frac{\mu_i}{d_i}\Psi_i^L(t) + (1-\mu_i)\phi_i(t)\phi_i(t)^\top \tag{11}$$

for $t \in \mathbb{N}$. Then, Ψ_i^L represents the unperturbed steady-state trajectory for Ψ (namely, the equivalent of $\Psi^\star$) in case the information coming from other agents is discarded in (3) (i.e. if $\sum_{k\in I_i}\Psi_k$ is substituted by Ψ_i). Therefore, the condition $\Psi_i^L(t) \ge \varepsilon_i I$ refers to an excitation property concerning only the local signal ϕ_i, hence called "local excitation". Lemma 8, whose proof is omitted for space reasons, establishes that local excitation always implies that agent i is excited in the sense of Definition 6 if the information coming from the neighboring agents is not discarded.

Lemma 8. *For every $i \in \mathcal{N}$, $t \in \mathbb{N}$ and $\varepsilon_i > 0$, we have $\Psi_i^L(t) \ge \varepsilon_i I \implies \Psi_i^\star(t) \ge \varepsilon_i I$.*

The converse implication is instead false in general. This is proved by the following trivial example: take $\mathcal{N} = \{1, 2\}$ ($n = 2$), $I_1 = I_2 = \{1, 2\}$, $\mu_1 = \mu_2 = 1/2$, $\phi_1(t) = (1, 0)$ and $\phi_2(t) = (0, 1)$ for all $t \in \mathbb{N}$. For both $i = 1, 2$, $\Psi_i^L(t)$ is singular for every $t \in \mathbb{N}$. Hence, local excitation does not hold for any agent. Instead, both agents are $(1/8, 2)$-excited according to Definition 6. Hence, in view of Lemma 8 and the previous example, we conclude that the excitation requirement in Definition 6 is *strictly weaker* than local excitation. Moreover, we remark that such excitation requirement may be weakened further if one uses regularization, as explained later in Section 5.2.

3.5. State-to-output stability and design of γ_i

The only degrees of freedom of (3)–(4) that remain to be fixed are the functions γ_i. In this section, we choose γ_i by following Bin (2022) in order to force a state-to-output stability property (made precise by Lemma 9) that will be needed for the main convergence result reported later in Section 4. At a first glance, Proposition 7 seems to suggest that a reasonable choice for $\gamma_i(\Psi_i, \eta_i)$ is $\gamma_i(\Psi_i, \eta_i) := \Psi_i^{-1}\eta_i$. This choice, however, is only valid at the ideal steady state where $\Psi_i = \Psi_i^\star$ and $\eta_i = \eta_i^\star$ and if agent i is (ε_i, τ_i)-excited for some (ε_i, τ_i). During the transitory, indeed, $\Psi_i(t)$ can be singular or of variable rank and, with such choice of γ_i, we would fail to establish a relation between the difference $\gamma_i(\Psi_i(t),\eta_i(t))-\gamma_i(\Psi_i^\star(t),\eta_i^\star(t))$ and the deviation $x(t)-x^\star(t)$ from the ideal steady state. Such a relation is needed to bound $|\hat\theta_i-\theta|$ by using the stability result of Proposition 3. To avoid these problems, in this section we use the following construction. With $\underline{\varepsilon}_i>0$ arbitrary and $\bar\phi$, $\bar y$ given by Assumption 1, define compact the set

$$\Gamma_i := \left\{(\Psi_i,\eta_i)\in\mathbb{P}_{n_\theta}\times\mathbb{R}^{n_\theta}\ :\ |\Psi_i|_\infty \le \left(\frac{1-\underline{\mu}}{1-\bar\mu}\right)\bar\phi^2+1\right.$$
$$\left. |\eta_i|_\infty \le \left(\frac{1-\underline{\mu}}{1-\bar\mu}\right)\bar\phi\bar y+1,\ \Psi_i\ge\frac{\underline{\varepsilon}_i}{2}I\right\}.$$

Then, we pick $\gamma_i:\mathbb{P}_{n_\theta}\times\mathbb{R}^{n_\theta}\to\mathbb{R}^{n_\theta}$ in such a way that:

G1. γ_i is continuous and bounded.
G2. $\gamma_i(\Psi_i,\eta_i)=\Psi_i^{-1}\eta_i$ for all $(\Psi_i,\eta_i)\in\Gamma_i$.

Then, by means of the same arguments used in Bin (2022, Lemma 1 and Lemma 4), the following property can be established.

Lemma 9. *For each $i\in\mathcal{N}$, let γ_i be chosen such that* ***G1*** *and* ***G2*** *hold. Then, there exists $\ell_i>0$ such that*

$$|\gamma_i(\Psi_i,\eta_i)-\gamma_i(\Psi_i',\eta_i')|_\infty \le \ell_i\Big(|\Psi_i-\Psi_i'|_\infty+|\eta_i-\eta_i'|_\infty\Big)$$

holds for all $(\Psi_i,\eta_i)\in\Gamma_i$ and all $(\Psi_i',\eta_i')\in\mathbb{P}_{n_\theta}\times\mathbb{R}^{n_\theta}$.

In view of Proposition 2, from (7) we obtain $|\Psi_i^\star(t)|_\infty\le|\Psi^\star(t)|_\infty\le(1-\underline{\mu})(1-\bar\mu)^{-1}\bar\phi^2$ and $|\eta_i^\star(t)|_\infty\le|\eta^\star(t)|_\infty\le(1-\underline{\mu})(1-\bar\mu)^{-1}\bar\phi\bar y$ for all $t\in\mathbb{N}$. Hence, if agent i is (ε_i,τ_i)-excited with $\varepsilon_i\ge\underline{\varepsilon}_i$, then $(\Phi_i^\star(t),\eta_i^\star(t))\in\Gamma_i$ for all $t\ge\tau_i$, and Lemma 9 implies that the state-to-output stability property

$$|\gamma_i(\Psi_i^\star(t),\eta_i^\star(t))-\gamma_i(\Psi_i(t),\eta_i(t))|_\infty\le 2\ell_i|x^\star(t)-x(t)| \tag{12}$$

holds for every solution pair $(x,(y,\phi,\delta))$ of (5) and for all $t\ge\tau_i$.

Remark 10 (*Choice of γ_i*). A possible choice of γ_i satisfying the above properties consists in taking $\gamma_i(\Psi_i,\eta_i)$ as a vector with kth component $\gamma_i(\Psi_i,\eta_i)_k := \mathrm{sat}_{\bar\theta_i}(u_{i,k})$, where $u_{i,k}$ is the kth component of $\Psi_i^\dagger\eta_i$ ($\cdot^\dagger$ denotes the Moore–Penrose pseudoinverse), $\bar\theta_i := 2(\bar\phi\bar y(1-\underline{\mu})(1-\bar\mu)^{-1}+1)/\underline{\varepsilon}_i$, and $\mathrm{sat}_\star(\cdot) := \min\{\max\{\cdot,-\star\},\star\}$ denotes the standard saturation function. See Bin (2022) for further details.

4. Main result

Let ω be the class-$\mathcal{K}$ function defined in (8) for which Proposition 3 holds and, for each $i\in\mathcal{N}$, let $\underline{\varepsilon}_i>0$ be the constant used in the construction of the set Γ_i and $\ell_i>0$ be such that Lemma 9 holds. Then, the following theorem establishes the main result of the paper, which relates the asymptotic estimation error on θ to the “asymptotic magnitude” of the disturbance δ.

Theorem 11. *Suppose that Assumption 1 holds. Then, every solution pair $(x,(y,\phi,\delta))$ of (5) with δ bounded is bounded. Moreover, if agent $i\in\mathcal{N}$ is $(\varepsilon_i,\tau_i^\star)$-excited for some $\tau_i\in\mathbb{N}$ and with $\varepsilon_i\ge\underline{\varepsilon}_i$, then*

$$\limsup_{t\to\infty}|\hat\theta_i(t)-\theta|_\infty\le 2\ell_i\omega\left(\limsup_{t\to\infty}|\delta(t)|_\infty\right). \tag{13}$$

Proof. In view of (7), under Assumption 1 $x^\star$ is bounded. Then, boundedness of x when δ is bounded follows by Proposition 3. Moreover, by means of standard ISS arguments (see, e.g. Cai and Teel (2009, Lemma 3.6)), one can deduce from Proposition 3 that

$$\limsup_{t\to\infty}|x(t)-x^\star(t)|\le\omega\left(\limsup_{t\to\infty}|\delta(t)|_\infty\right). \tag{14}$$

Finally, if agent i is (ε_i,τ_i)-excited with $\varepsilon_i\ge\underline{\varepsilon}_i$ then, as shown at the end of Section 3.5, $(\Psi_i^\star(t),\eta_i^\star(t))\in\Gamma_i$ for all $t\ge\tau_i$. Since in this case Proposition 7 implies $\theta=\Psi_i^\star(t)^{-1}\eta_i^\star(t)=\gamma_i(\Psi_i^\star(t),\eta_i^\star(t))$ for $t\ge\tau_i$, then the claim follows from (12) and (14).

As a direct corollary of Theorem 11 we obtain that, if every agent $i\in\mathcal{N}$ is $(\varepsilon_i,\tau_i^\star)$-excited for some $\tau_i\in\mathbb{N}$ and with $\varepsilon_i\ge\underline{\varepsilon}_i$, then (13) holds for all $i\in\mathcal{N}$. Hence, in this case, the robust estimation problem introduced in Section 2.3 is solved.

5. Convergence, regularization and scalability

5.1. Exact convergence and convergence rate

For given $i\in\mathcal{N}$, δ and (ϕ,y) satisfying Assumption 1, we say that the limit (13) holds uniformly if the following property holds: for every $\epsilon>0$ and $r>0$, there exists $\bar t>0$, such that every solution pair $(x,(y,\phi,\delta))$ with $|x(0)|\le r$ satisfies $|\hat\theta_i(t)-\theta|_\infty\le 2\ell_i\omega(\limsup_{t\to\infty}|\delta(t)|_\infty)+\epsilon$ for all $t\ge\bar t$. Theorem 11 and Proposition 3 guarantee that, if agent $i\in\mathcal{N}$ is $(\varepsilon_i,\tau_i^\star)$-excited for some $\tau_i\in\mathbb{N}$ and with $\varepsilon_i\ge\underline{\varepsilon}_i$, then (13) is uniform. Moreover, we stress that, if $\lim_{t\to\infty}\delta(t)=0$, then (13) can be strengthened to $\lim_{t\to\infty}\hat\theta_i(t)=\theta$, which is exact convergence, and convergence is exponential. Finally, we remark that, in this unperturbed case, the excitation condition of Definition 6 is only a sufficient condition for convergence.

5.2. On the use of regularizers

As the excitation condition of Definition 6 cannot be checked a priori in general, one may wonder whether such assumption can be weakened. It turns out it is possible at the price, however, of introducing a bias in the estimates. In particular, we can pick arbitrarily a matrix $\Omega_i\in\mathbb{P}_{n_\theta}$ satisfying $\Omega_i\ge\underline{\varepsilon}_iI$ and then choose γ_i such that the following property holds in place of Item **G2**:

G2’. $\gamma_i(\Psi_i,\eta_i)=(\Psi_i+\Omega_i)^{-1}\eta_i$ for all $(\Psi_i,\eta_i)\in\Gamma_i'$,

in which Γ_i' is defined by modifying Γ_i as follows

$$\Gamma_i' := \left\{(\Psi_i,\eta_i)\in\mathbb{P}_{n_\theta}\times\mathbb{R}^{n_\theta}\ :\ |\eta_i|_\infty\le\left(\frac{1-\underline{\mu}}{1-\bar\mu}\right)\bar\phi\bar y+1\right.$$
$$\left. |\Psi_i|_\infty\le\left(\frac{1-\underline{\mu}}{1-\bar\mu}\right)\bar\phi^2+|\Omega_i|_\infty+1,\ \Psi_i\ge\frac{\underline{\varepsilon}_i}{2}I\right\}.$$

Since $\Psi_i^\star(t)\in\mathbb{P}_{n_\theta}$, then $\Psi_i^\star(t)+\Omega_i\ge\underline{\varepsilon}_iI$ holds for all $t\in\mathbb{N}$ and, hence, $(\Psi_i^\star(t)+\Omega_i,\eta_i^\star(t))\in\Gamma_i'$ for all $t\in\mathbb{N}$. Therefore, the excitation condition of Definition 6 is not needed anymore to conclude (12). Furthermore, in this case also the saturation of γ_i (Item **G1**) can be avoided. Indeed, since $\Psi_i(t)\in\mathbb{P}_{n_\theta}$ for all $t\in\mathbb{N}$ and all solution pairs $(x,(y,\phi,\delta))$ of (5), then $\Omega_i\ge\varepsilon_iI$ also implies $\Psi_i(t)+\Omega_i\ge\underline{\varepsilon}_iI$ for all $t\in\mathbb{N}$. Hence, by using the fact that, for every $A,B\in\mathbb{P}_n$ satisfying $A,B\ge\varepsilon I$, and for every $x,y\in\mathbb{R}^{n_\theta}$, one has $|A^{-1}x-B^{-1}y|_\infty=|(A^{-1}-B^{-1})x+B^{-1}(x-y)|_\infty\le(n/\varepsilon^2)|x|_\infty|A-B|_\infty+(\sqrt{n}/\varepsilon)|x-y|_\infty$, we obtain

$$|(\Psi_i^\star(t)+\Omega_i)^{-1}\eta_i^\star(t)-(\Psi_i(t)+\Omega_i)^{-1}\eta_i(t)|_\infty$$
$$\le\frac{n_\theta(1-\underline{\mu})\bar\phi\bar y}{(1-\bar\mu)\varepsilon^2}|\Psi_i^\star(t)-\Psi_i(t)|_\infty+\frac{\sqrt{n_\theta}}{\varepsilon}|\eta_i^\star(t)-\eta_i(t)|_\infty,$$

which implies (12) with $2\ell_i=n_\theta(1-\underline{\mu})\bar\phi\bar y(1-\bar\mu)^{-1}\varepsilon^{-2}+\sqrt{n_\theta}\varepsilon-1$. Thus, if $\Omega_i\ge\varepsilon_iI$, one can simply pick

$$\gamma_i(\Psi_i,\eta_i):=(\Psi_i+\Omega_i)^{-1}\eta_i. \tag{15}$$

Nevertheless, if $\Omega_i\neq0$, one cannot conclude (13) anymore. Instead, one obtains the weaker inequality

$$\limsup_{t\to\infty}|\hat\theta_i(t)-\theta|_\infty\le c_i(\Omega_i)+2\ell_i\omega\left(\limsup_{t\to\infty}|\delta(t)|_\infty\right),$$

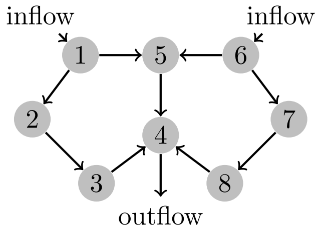


Fig. 1. Road network.

in which $c_i(\Omega_i) \geq 0$ is a bias term (see Bin, 2022, Section 6.1 for further details).

5.3. Remarks on scalability and decentralization

The update laws (3) are decentralized, since the update of the state variables of each agent i only depends on the state variables of the neighboring agents $k \in I_i$. Moreover, the "gains" μ_i are arbitrary, each agent can fix $\mu_i \in (0, 1)$ independently from the others. The construction of γ_i, however, uses centralized quantities since the set Γ_i relies on the knowledge of $\underline{\mu}$, $\bar{\mu}$ and the constants $\bar{\phi}$ and $\bar{y}$ given by Assumption 1. Nevertheless, we observe that: **(i)** The quantities $\underline{\mu}$, $\bar{\mu}$, $\bar{\phi}$, and $\bar{y}$ are only used to estimate an a priori upper bound on the norm of $\Psi_i^\star$ and $\eta_i^\star$. If such quantities are not available, an agent can either estimate the upper bound by looking at the local quantities $|\Psi_i(t)|_\infty$ and $|\eta_i(t)|_\infty$, or take a very large upper bound that is likely to be achieved. **(ii)** If agent i uses a regularization matrix $\Omega_i \geq \varepsilon_i I$ then, as explained in previous Section 5.2, γ_i can be chosen as in (15), and thus no knowledge of $\underline{\mu}$, $\bar{\mu}$, $\bar{\phi}$, and $\bar{y}$ is required anymore. Thus, in this case each agent can design γ_i independently from the other agents and centralized quantities.

Overall, we can therefore conclude that (3) has good decentralization and scalability properties for what concerns the definition of the update laws. Finally, we observe that the function ω (defined in (8)), for which the claim of Proposition 3 holds, is proportional to the factor $(1-\underline{\mu})(1-\bar{\mu})^{-1}$. Thus, while the choice of each gain μ_i is independent from the others, the more similar the gains are, the better it is from the standpoint of sensitivity to disturbances. In the limit case in which the gains satisfy $\mu_1 = \cdots = \mu_n$, then $(1-\underline{\mu})(1-\bar{\mu})^{-1} = 1$, and ω does not depend on them.

6. Application to road pricing

Given a road network, represented by a directed graph $(\mathcal{V}, \mathcal{N})$ as in Fig. 1, with $\mathcal{V}$ a set of vertices representing crossings and $\mathcal{N} \subset \mathcal{V}^2$ the set of roads, the problem of road pricing pertains the assignment of a toll τ_i to each road $i \in \mathcal{N}$ with the aim of mitigating congestion. Road pricing boasts a long academic history, especially in the economics community (Pigou, 1920; Small & Verhoef, 2007; Yang & Huang, 2005). The basic idea follows the principle of *marginal cost* (or Pigovian tax): Efficiency is obtained when each driver pays a toll balancing the externalities caused by their journey. Formally, if d_i denotes the density of vehicles on road $i \in \mathcal{N}$, and $\ell(d_i)$ the corresponding *latency* (the average travel time), then the marginal cost pricing theory suggests taking

$$\tau_i = \ell'(d_i)d_i, \tag{16}$$

where ℓ' denotes the derivative of ℓ (Yang & Huang, 2005, sec. 3.2) (here we are assuming that the latency function ℓ is the same for each road, assumption that is justified when considering similar roads as we do in the forthcoming simulations). In this section, we consider the problem of *adaptive decentralized marginal cost pricing* described hereafter. We assume that with each arc i there is associated a unique agent (labeled by i as the corresponding arc). Agents aim to implement the marginal cost policy (16) in a distributed way, with each agent responsible of deciding the toll levied on its arc. To implement (16), agents need the function ℓ', which is unknown a priori (the uncertain form of such functions is, indeed, one of the main obstacles for the implementation of marginal cost pricing; see, e.g., Yang, Meng, & Lee, 2004). Hence, ℓ' must be estimated at run time. We approach this problem by applying the methodology developed in the previous sections. Specifically, we suppose that each agent i can measure (with the due approximation) the density $d_i(t)$ and the corresponding latency $\ell(d_i(t))$ on the associated road i at each time t. In terms of (2), we thus have $y_i(t) = \ell(d_i(t))$ and $\phi_i(t) = (\psi_1(d_i(t)), \dots, \psi_{n_\theta}(d_i(t)))$ for some arbitrary $\mathcal{C}^1$ basis-functions ψ_k. Furthermore, we assume that agents can communicate with those associated with neighboring arcs. Namely, we assume that an information network (Section 2.2) $\mathcal{C} = \{I_i\}_{i\in\mathcal{N}}$ is given with $I_i := \{j \in \mathcal{N} : j_1 = i_1 \vee j_1 = i_2 \vee j_2 = i_1 \vee j_2 = i_2\}$ for all $i \in \mathcal{N}$, where, for an arc $k \in \mathcal{N}$, we let $k_1, k_2 \in \mathcal{V}$ be such that $k = (k_1, k_2)$. In this setting, we consider the approximation $\ell \approx \sum_{k=1}^{n_\theta} \theta_k \psi_k$ and we equip each agent with the distributed estimation scheme (3)–(4) to find out the best parameters $\theta_1, \dots, \theta_{n_\theta}$ from data. Specifically, with $\hat{\theta}_i$ denoting the approximation of $\theta := (\theta_1, \dots, \theta_{n_\theta})$ that agent i obtains from (4), each agent approximates ℓ' by $\widehat{\ell'}_i := \hat{\theta}_i^\top \psi'$, where $\psi' := (\psi_1', \dots, \psi_{n_\theta}')$. A similar problem has been considered in Poveda, Brown, Marden, and Teel (2017) where, however, each agent estimates its own parameters alone without communicating with others. We stress that, in our context, estimating the parameters θ in a collective way has the advantage that each agent can exploit the information coming from the traffic data in other roads. This is particularly useful for agents controlling empty or low-congested roads, as otherwise they could not obtain a meaningful estimate from their own measurements and thus levy a proper toll (this is formally captured by the difference between excitation according to Definition 6 and local excitation; see the discussion in Section 3.4). In the forthcoming simulation, we consider a road network represented by the graph shown in Fig. 1. The simulation setting is the following. Arcs represent one-way single-lane paths. Their length L_i is measured in cells (a cell is the discrete unit of space) and, in the following, it is set to $L_i = 50$ for all $i \in \mathcal{N}$. Vehicles flows are simulated microscopically and in discrete-time. Each vehicle v enters the network with a given origin $O_v \in \mathcal{V}$ and destination $D_v \in \mathcal{V}$, and it has state variable $(\alpha_v, c_v) \in \mathcal{N} \times \mathbb{N}$, where α_v represents the current arc and c_v the current cell occupied by v on α_v. These variables are updated as

$$\alpha_v^+ = \begin{cases} \alpha_v & \text{if } c_v < L_{\alpha_v} \text{ or } (J(\alpha_v), 1) \text{ is occupied} \\ J(\alpha_v) & \text{otherwise} \end{cases}$$

$$c_v^+ = \begin{cases} c_v & \text{if } c_v = L_{\alpha_v} \text{ and } (J(\alpha_v), 1) \text{ is occupied} \\ 1 & \text{if } c_v = L_{\alpha_v} \text{ and } (J(\alpha_v), 1) \text{ is free} \\ c_v + K_v & \text{otherwise} \end{cases}$$

in which $K_v := \sup\{k \in \mathbb{N}_{\leq V} : c_v + k \leq L_{\alpha_v} \text{ and } (\alpha_v, c_v + h) \text{ is not occupied for all } h \in \mathbb{N}_{\leq k}\}$ denotes the maximal number of cells that v can advance, where $V = 4$ denotes a common maximal speed (in cells per time units), and $J(\alpha_v)$ is the first arc produced by a shortest-path algorithm finding the shortest weighted path[1] between the current node to which α_v is incident and the vehicle's destination D_v. Here, the weights on each $i \in \mathcal{N}$ at each $t \in \mathbb{N}$ are given by $w_i(t) := L_i/V + \tau_i(t)$, namely, by the sum of an

[1] This corresponds to the simplifying (yet widespread) assumption of fully rational drivers with perfect knowledge of the network weights.

expected travel time L_i/V when no congestion is present, and the time-equivalent of the toll levied on i at time t (here, we assume that all vehicles have a unitary value of time factor, so as we can measure $\tau_i(t)$ in time units). The order of update of the vehicles is chosen randomly at each time. Vehicles are removed once they reach their destination. We denote by $\mathcal{H}^t$ the set of all vehicles on the road at time t. For the estimation phase, we pick $n_\theta = 6$ and choose the basis functions $\psi_1, \dots, \psi_6$ as the elements of a biorthogonal spline basis[2] for $L_2([0, 1])$ (Daubechies, 1992, pp. 271–280). Each agent is then equipped with the estimator (3)–(4). For simplicity, the update of the estimation law is synchronized with that of the vehicles. The density $d_i(t)$ and latency $y_i(t) = \ell(d_i(t))$ on road i are estimated by each node from microscopic observations as $d_i(t) \approx \frac{\operatorname{card}\mathcal{H}_i^t}{L_i}$ and $y_i(t) \approx L_i \cdot (\frac{1}{\operatorname{card}\mathcal{H}_i^t} \sum_{v\in\mathcal{H}_i^t} s_v^t)^\dagger$, where $\mathcal{H}_i^t := \{v \in \mathcal{H}^t \ :\ \alpha_v(t) = i\}$ is the set of vehicles on road i at time t, and s_v^t is the speed of vehicle v at time t, estimated as $s_v^t = c_v(t) - c_v(t-1)$ if $\alpha_v(t) = \alpha_v(t-1)$, or $s_v^t = 1$ otherwise. The design of the parameters in (3)–(4) is made in a fully decentralized way. In particular, the parameters μ_i are chosen as $\mu_i = 0.99 + i_1 \cdot 10^{-3}$ (i_1 denotes the tail node of arc i). Moreover, for each i, γ_i is chosen as in (15) with $\Omega_i = 10^{-3}I$. We stress that the usage of the above-defined approximations of $d_i(t)$ and $y_i(t)$, which are quite rough macroscopic approximations obtained by averaging microscopic quantities, is only theoretically justified thanks to the robustness properties guaranteed by Proposition 3 and Theorem 11, which ensure that the employed least squares algorithm remains stable for any bounded uncertainty in $\ell(d_i(t))$, $d_i(t)$, and in the approximation of ℓ. Figs. 2–5 show the results of a simulation running for $T = 1000$ units of time with the inflow from node 1 to node 4 at full capacity and that from 6 to 4 at 80%. Namely, at each time t, after the update of the vehicles' state variables, if the first cell of the first arc of the current shortest path connecting 1 and 4 (resp. 6 and 4) is free, then with probability 1 (resp. 0.8) a new vehicle v with origin–destination pair $(O_v, D_v) = (1, 4)$ (resp. $(O_v, D_v) = (6, 4)$) is added to the network. In particular, Fig. 2 shows that, without tolls, all drivers seek the path that would be the shortest in absence of congestion. Consequently, we observe a high congestion concentrated in few roads, which provokes a low traveling speed. Instead, Fig. 3 shows a simulation where tolls are levied according to the adaptive methodology described previously. As the figure clearly shows, the levied tolls have the effect of distributing more equally the drivers over the networks' roads, with the consequence of a lower congestion on each road and a larger mean speed. Fig. 4 shows a comparison between these two simulations in terms of mean congestion, mean speed, and mean travel time from the origin to the destinations. As shown by the figure, the developed adaptive pricing mechanism permits to more than double the mean speed while more than halving congestion and travel time. Finally, for the case where tolls are levied, Fig. 5 shows the dynamics of the estimated parameters. As it can be seen from the figure, all agents estimate similar parameters as expected.

7. Conclusions

We proposed a distributed recursive least squares algorithm for the estimation of an unknown parameter over a network. With respect to the other distributed approaches, the main novelty of this work consists in the robustness property to general bounded disturbances in all the variables, characterized in terms of ISS. In addition, linear convergence is shown under a cooperative excitation condition weaker than local excitation. Scalability properties were also discussed, and the methodology was validated on a road pricing example.

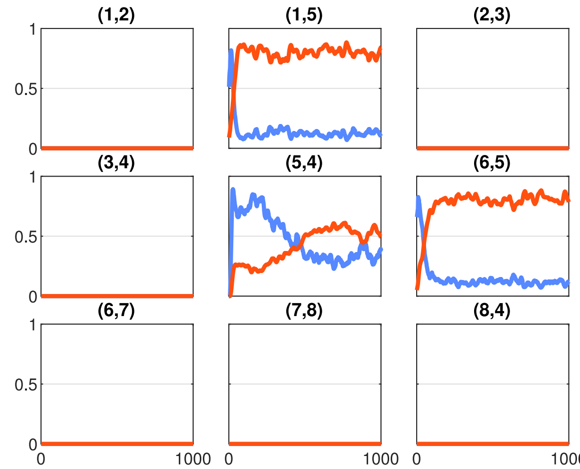


Fig. 2. Time series of the congestion (red) and mean speed (blue) on each road in absence of tolls. The speed is normalized with respect to the maximum speed V. The time series are averaged on a moving window of magnitude 15 time units. Label (i, j) denotes the arc (road) from node i to node j in the graph of Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

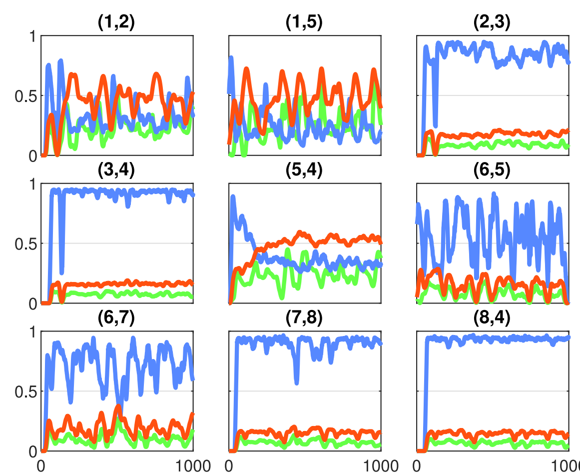


Fig. 3. Time series of the congestion (red), mean speed (blue), and levied toll (green) on each road in presence of tolls. The speed is normalized with respect to the maximum speed V, the tolls are normalized with respect to the maximum toll levied. The time series are averaged on a moving window of magnitude 15 time units. Label (i, j) denotes the arc (road) from node i to node j in the graph of Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

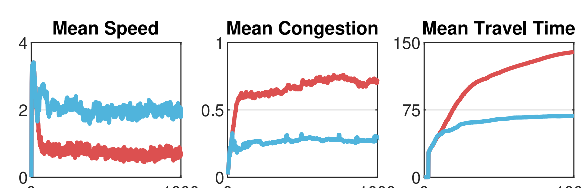


Fig. 4. Comparison between performances when no tolls are levied (red) and when tolls are levied (blue). The mean speed and the mean congestion are computed by averaging the corresponding time series shown in Figs. 2 and 3 over the roads with non-zero occupation. The mean travel time is the average of the time each driver takes to go from its origin to its destination. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[2] Specifically, ψ_k can be obtained in MATLAB as $\psi_k(\cdot) := 2^{-s/2}\varphi(2^{-s}\cdot - k)$ with $s = -2$ and where φ is the dual scaling function obtained with the command *wavefun* of the Wavelet Toolbox with argument 'bior3.5'.

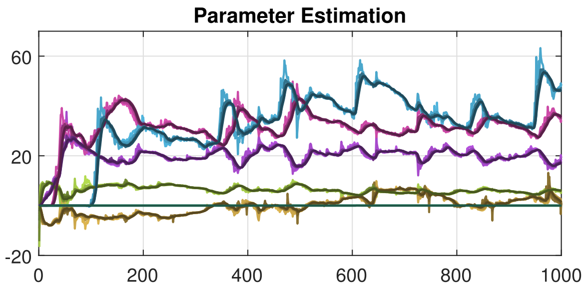


Fig. 5. Estimated parameters when tolls are levied. Brighter and darker shades of the same color are used to plot the time series of the same parameter for different agents.

Appendix A. Proof of Proposition 3

Pick arbitrarily a solution pair $(x, (y, \phi, \delta))$ to (5), and let $x^\star$ be defined as in (7). By direct solution, (5) yields

$$\Psi(t) = F^t\Psi(0) + \sum_{s=0}^{t-1} F^{t-s-1}G\Phi(\phi(s) + \delta_\phi(s))$$

for all $t \in \mathbb{N}$. Define $\widetilde{\Psi} := \Psi - \Psi^\star$ and $\widetilde{\Phi}(\phi, \delta_\phi) := \Phi(\phi + \delta_\phi) - \Phi(\phi)$. In view of (7), we obtain

$$|\widetilde{\Psi}(t)|_\infty \le \bar{\mu}^t|\widetilde{\Psi}(0)|_\infty + (1 - \underline{\mu})\sum_{s=0}^{t-1} \bar{\mu}^{t-s-1}|\widetilde{\Phi}(\phi(s), \delta_\phi(s))|_\infty,$$

for all $t \in \mathbb{N}$, where we have used the fact that, by construction, $\Psi^\star(0) = 0$, $|G|_\infty = 1 - \underline{\mu}$, and that, in view of Proposition 2, $|F|_\infty \le \bar{\mu}$. Similar arguments show that, with $\tilde{\eta} := \eta - \eta^\star$ and $\widetilde{\Xi}(\phi, y, d) := \Xi(\phi + \delta_\phi, y + \delta_y) - \Xi(\phi, y)$, we have

$$|\tilde{\eta}(t)|_\infty \le \bar{\mu}^t|\tilde{\eta}(0)|_\infty + (1 - \underline{\mu})\sum_{s=0}^{t-1} \bar{\mu}^{t-s-1}|\widetilde{\Xi}(\phi(s), y(s), \delta(s))|_\infty$$

for all $t \in \mathbb{N}$. Next, observe that, for all $a, b \in \mathbb{R}^{n_\theta}$ and $c, d \in \mathbb{R}$, $|aa^\top - bb^\top|_\infty \le 2|a|_\infty|a - b|_\infty + |a - b|_\infty^2$ and $|ac - bd|_\infty \le |a|_\infty|c - d|_\infty + |c|_\infty|a - b|_\infty + |a - b|_\infty|c - d|_\infty$. By using these relations with $a = \phi(s)$, $b = \phi(s) + \delta_\phi(s)$, $c = y(s)$, and $d = y(s) + \delta_y(s)$, in view of Assumption 1 we obtain $\max\{|\widetilde{\Phi}(\phi(s), \delta_\phi(s))|_\infty, |\widetilde{\Xi}(\phi(s), y(s), \delta(s))|_\infty\} \le 2\max\{\bar{\phi}, \bar{y}\}|\delta(s)| + |\delta(s)|^2$ for all $s \in \mathbb{N}$. Hence, $\tilde{x} := x - x^\star$ satisfies

$$\begin{aligned}|\tilde{x}(t)| &= \max\left\{|\widetilde{\Psi}(t)|_\infty, |\tilde{\eta}(t)|_\infty\right\} \\ &\le \bar{\mu}^t|\tilde{x}(0)| + \left(\frac{1 - \underline{\mu}}{1 - \bar{\mu}}\right)\left(2\max\{\bar{\phi}, \bar{y}\}|\delta|_{\infty,t} + |\delta|_{\infty,t}^2\right),\end{aligned}$$

where we let $|\delta|_{\infty,t} := \sup_{s\in\mathbb{N}_{\le t-1}} |\delta(s)|_\infty$ and we used the fact that $\sum_{s=0}^{t} \bar{\mu}^{t-s-1} \le \sum_{s=0}^{\infty} \bar{\mu}^s = (1 - \bar{\mu})^{-1}$. □

Appendix B. Proof of Lemma 5

For every $t \in \mathbb{N}$, we have $\Psi_i^\star(t)\theta = \eta_i^\star(t)$ for all $i \in \mathcal{N}$ if and only if

$$\Psi^\star(t)\theta := (\Psi_i^\star(t)\theta)_{i\in\mathcal{N}} = \eta^\star(t). \tag{B.1}$$

It thus suffices to prove that (B.1) holds for all $t \in \mathbb{N}$. We proceed by induction. Suppose that at some $t \in \mathbb{N}$, (B.1) holds. Then, at $t + 1$ we have $\Psi^\star(t + 1)\theta - \eta^\star(t + 1) = F(\Psi^\star(t)\theta - \eta^\star(t)) + G\big(\Phi(\phi(t))\theta - \Xi(\phi(t), y(t))\big) = G\big(\Phi(\phi(t))\theta - \Xi(\phi(t), y(t))\big)$. By definition of Φ and Ξ, and using (1), we obtain $\Phi(\phi(t))\theta := (\phi_i(t)\phi_i(t)^\top\theta)_{i\in\mathcal{N}} = (\phi_i(t)y_i(t))_{i\in\mathcal{N}} = \Xi(\phi(t), y(t))$, and thus we conclude that $\Phi(t+1)\theta - \eta(t+1) = 0$, which implies $\Psi^\star(t+1)\theta = \eta^\star(t+1)$. Since $\Psi^\star(0) = 0$ and $\eta^\star(0) = 0$, then (B.1) holds at $t = 0$. Therefore, the claim follows by induction on t. □

References

Arablouei, R., Doğançay, K., Werner, S., & Huang, Y.-F. (2014). Adaptive distributed estimation based on recursive least-squares and partial diffusion. *IEEE Transactions on Signal Processing*, *62*(14), 3510–3522.

Bin, M. (2022). Generalized recursive least squares: Stability, robustness, and excitation. *Systems & Control Letters*, [ISSN: 0167-6911] *161*, Article 105144.

Breschi, V., Bemporad, A., & Kolmanovsky, I. V. (2020). Cooperative constrained parameter estimation by ADMM–RLS. *Automatica*, *121*, Article 109175.

Cai, C., & Teel, A. R. (2009). Characterizations of input-to-state stability for hybrid systems. *Systems & Control Letters*, *58*(1), 47–53.

Cattivelli, F. S., Lopes, C. G., & Sayed, A. H. (2008). Diffusion recursive least-squares for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, *56*(5), 1865–1877.

Cattivelli, F. S., & Sayed, A. H. (2010). Analysis of spatial and incremental LMS processing for distributed estimation. *IEEE Transactions on Signal Processing*, *59*(4), 1465–1480.

Chen, W., Hua, S., & Sam Ge, S. (2014). Consensus-based distributed cooperative learning control for a group of discrete-time nonlinear multi-agent systems using neural networks. *Automatica*, *50*(9), 2254–2268.

Chen, J., & Sayed, A. H. (2012). Diffusion adaptation strategies for distributed optimization and learning over networks. *IEEE Transactions on Signal Processing*, *60*(8), 4289–4305.

Chen, W., Wen, C., Hua, S., & Sun, C. (2014). Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative PE condition via consensus. *IEEE Transactions on Automatic Control*, *59*(1), 91–106.

Daubechies, I. (1992). *Ten lectures on wavelets*. SIAM.

Javed, M. U., Poveda, J. I., & Chen, X. (2022). Excitation conditions for uniform exponential stability of the cooperative gradient algorithm over weakly connected digraphs. *IEEE Control Systems Letters*, [ISSN: 2475-1456] *6*, 67–72.

Jiang, Z.-P., Teel, A. R., & Praly, L. (1994). Small-gain theorem for ISS systems and applications. *Mathematics of Control, Signals, and Systems*, *7*(2), 95–120.

Johansson, B., Keviczky, T., Johansson, M., & Johansson, K. H. (2008). Subgradient methods and consensus algorithms for solving convex optimization problems. In *2008 47th IEEE conf. decis. control* (pp. 4185–4190).

Li, L., Zhao, H., & Lv, S. (2021). Diffusion recursive total least square algorithm over adaptive networks and performance analysis. *Signal Processing*, *182*, Article 107954.

Liu, Z., Liu, Y., & Li, C. (2014). Distributed sparse recursive least-squares over networks. *IEEE Transactions on Signal Processing*, *62*(6), 1386–1395.

Lopes, C. G., & Sayed, A. H. (2007). Incremental adaptive strategies over distributed networks. *IEEE Transactions on Signal Processing*, *55*(8), 4064–4077.

Lopes, C. G., & Sayed, A. H. (2008). Diffusion least-mean squares over adaptive networks: Formulation and performance analysis. *IEEE Transactions on Signal Processing*, *56*(7), 3122–3136.

Mateos, G., & Giannakis, G. B. (2012). Distributed recursive least-squares: Stability and performance analysis. *IEEE Transactions on Signal Processing*, *60*(7), 3740–3754.

Mateos, G., Schizas, I. D., & Giannakis, G. B. (2009). Distributed recursive least-squares for consensus-based in-network adaptive estimation. *IEEE Transactions on Signal Processing*, *57*(11), 4583–4588.

Pigou, A. C. (1920). *The economics of welfare*. London: Macmillan and Co..

Poveda, J. I., Brown, P. N., Marden, J. R., & Teel, A. R. (2017). A class of distributed adaptive pricing mechanisms for societal systems with limited information. In *2017 IEEE 56th conf. decis. control* (pp. 1490–1495).

Rastegarnia, A. (2019). Reduced-communication diffusion RLS for distributed estimation over multi-agent networks. *IEEE Transactions on Circuits and Systems II*, *67*(1), 177–181.

Sayed, A. H., Tu, S.-Y., Chen, J., Zhao, X., & Towfic, Z. J. (2013). Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior. *IEEE Signal Processing Magazine*, *30*(3), 155–171.

Schizas, I. D., Mateos, G., & Giannakis, G. B. (2009). Distributed LMS for consensus-based in-network adaptive processing. *IEEE Transactions on Signal Processing*, *57*(6), 2365–2382.

Small, K. A., & Verhoef, E. T. (2007). *The economics of urban transportation*. New York: Routledge.

Söderström, T. (2007). Errors-in-variables methods in system identification. *Automatica*, *43*, 939–958.

Sontag, E. D., & Wang, Y. (1996). New characterizations of input-to-state stability. *IEEE Transactions on Automatic Control*, *41*(9), 1283–1294.

Tu, S.-Y., & Sayed, A. H. (2012). Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, *60*(12), 6217–6234.

Wang, Z., Yu, Z., Ling, Q., Berberidis, D., & Giannakis, G. B. (2018). Decentralized RLS with data-adaptive censoring for regressions over large-scale networks. *IEEE Transactions on Signal Processing*, *66*(6), 1634–1648.

Xie, S., & Guo, L. (2018a). Analysis of distributed adaptive filters based on diffusion strategies over sensor networks. *IEEE Transactions on Automatic Control*, *63*(11), 3643–3658.

Xie, S., & Guo, L. (2018b). Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition. *SIAM Journal on Control and Optimization*, *56*(5), 3404–3431.

Xie, S., & Guo, L. (2018c). A necessary and sufficient condition for stability of LMS-based consensus adaptive filters. *Automatica*, *93*, 12–19.

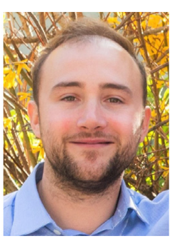
Xie, S., Zhang, Y., & Guo, L. (2021). Convergence of a Distributed Least Squares. *IEEE Transactions on Automatic Control*, *66*(10), 4952–4959.

Yang, H., & Huang, H. (2005). *Mathematical and economic theory of road pricing*. Amsterdam; Boston: Elsevier.

Yang, H., Meng, Q., & Lee, D.-H. (2004). Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions. *Transportation Research Part B: Methodological*, *38*(6), 477–493.

Yu, Y., Zhao, H., de Lamare, R. C., Zakharov, Y., & Lu, L. (2019). Robust distributed diffusion recursive least squares algorithms with side information for adaptive networks. *IEEE Transactions on Signal Processing*, *67*(6), 1566–1581.

Ilario Antonio Azzollini received the B.Sc. degree in Automation Engineering from the University of Bologna, Italy, July 2016, the M.Sc. degree Cum Laude in Systems and Control from the Delft University of Technology, The Netherlands, October 2018, and the Ph.D. degree in Control Theory from the University of Bologna, June 2022. Since November 2021 he is a R&D Software Engineer at the SACMI Innovation Lab, Italy. His research interests include nonlinear and adaptive control, distributed control, nonlinear optimization, automation and simulation software design, and control of mobile robots.



Michelangelo Bin received the Ph.D. degree in Control Theory in 2019 from the University of Bologna, Italy. Since October 2022 he is with the Department of Electrical, Electronic and Information Engineering, University of Bologna, Italy. Before he was Research Associate at Imperial College London, UK. He serves as Associate Editor of Systems & Control Letters. His research interests include systems theory, nonlinear control and regulation, and adaptive systems.

Lorenzo Marconi graduated in 1995 in electrical engineering from the University of Bologna. Since 1995 he has been with the Department of Electronics, Computer Science and Systems at University of Bologna, where he obtained his Ph.D. degree in 1998. From 1999 he has been an assistant professor in the same department where he is now a full professor. He has held visiting positions at various academic/research international institutions. He is a co-author of more than 250 technical publications on the subject of linear and nonlinear feedback designs published on international journals, books and conference proceedings. In 2005, he received the "Outstanding Application Paper Award" from IFAC for a coauthored paper published on Automatica. He is also the co-recipient of the 2014 IEEE Control Systems Magazine Outstanding Paper Award for the best paper published in the magazine in the period 2012–2013 and of the 2018 O. Hugo Schuck Best Paper Award. He is Fellow of IEEE for "contributions to feedback design of nonlinear systems and unmanned aerial vehicles". His current research interests include nonlinear control, output regulation, control of autonomous vehicles, fault detection and isolation, and fault tolerant control.

Thomas Parisini received the Ph.D. degree in Electronic Engineering and Computer Science in 1993 from the University of Genoa. He was with Politecnico di Milano and he currently holds the Chair of Industrial Control and serves as the Head of the Control and Power Research Group at Imperial College London. He is a Deputy Director of the KIOS Research and Innovation Centre of Excellence, University of Cyprus. Since 2001 he is also Danieli Endowed Chair of Automation Engineering with University of Trieste. In 2009–2012 he was Deputy Rector of University of Trieste. In 2018 he received an Honorary Doctorate from University of Aalborg, Denmark. He held a "Scholar-in-Residence" visiting position at Digital Futures-KTH, Stockholm, Sweden. He authored or co-authored a research monograph in the Communication and Control Series, Springer Nature, and over 400 research papers in archival journals, book chapters, and international conference proceedings. He is a co-recipient of the IFAC Best Application Paper Prize of the Journal of Process Control, Elsevier, for the three-year period 2011–2013 and of the 2004 Outstanding Paper Award of the IEEE Trans. on Neural Networks. He is also a recipient of the 2007 IEEE Distinguished Member Award. In 2016, he was awarded as Principal Investigator at Imperial of the H2020 European Union flagship Teaming Project KIOS Research and Innovation Centre of Excellence led by University of Cyprus. Thomas Parisini has served as the 2021–2022 President of the IEEE Control Systems Society and during 2009–2016 he was the Editor-in-Chief of the IEEE Trans. on Control Systems Technology. Since 2017, he is Editor for Control Applications of Automatica and since 2018 he is the Editor in Chief of the European Journal of Control. Among other activities, he was the Program Chair of the 2008 IEEE Conference on Decision and Control and General Co-Chair of the 2013 IEEE Conference on Decision and Control. Prof. Parisini is a Fellow of the IEEE and of the IFAC.