Automatica 121 (2020) 109175

Contents lists available at ScienceDirect

Automatica

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

Cooperative constrained parameter estimation by ADMM-RLS[☆]

Valentina Breschi [a,*], Alberto Bemporad [b], Ilya V. Kolmanovsky [c]

[a] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy
[b] IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy
[c] Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 26 March 2019
Received in revised form 16 June 2020
Accepted 4 July 2020
Available online xxxx

Keywords:
Estimation algorithms
Collaborative systems
Cloud computing
Iterative methods
Alternating Direction Method of Multipliers
Recursive least squares

ABSTRACT

With recent advances in cloud computing, resources with customizable computational power and memory can be exploited to store and analyze data collected from large sets of devices. Although one can exploit the connection to the cloud to perform all the desired tasks on the cloud itself, in many applications it is also desirable to retrieve and process information locally. In this paper, we present a collection of cloud-aided consensus-based Recursive Least-Squares (RLS) estimators. The approaches are tailored to handle linear and nonlinear consensus constraints and limitations on parameter ranges. All the methods are designed so that raw measurements collected at the device level are processed by the device itself, requiring minimal changes to (possibly pre-existing) RLS estimators. The local estimates are then recursively refined and fused on the cloud to reach consensus among the devices.

1. Introduction

The increasing connectivity of consumer devices allows both users and manufacturers to access and store data that can be used for various purposes, such as estimation, diagnostics and prognostics. Within this pervasively connected framework, devices can be seen as nodes of a network, that might cover a wide geographic area. This has stimulated research on distributed solutions for a variety of problems, such as fault detection (Boem, Ferrari, Keliris, Parisini, & Polycarpou, 2017), classification (Forero, Cano, & Giannakis, 2010), state (Olfati-Saber, 2007) and parameter (Mateos, Schizas, & Giannakis, 2009) estimation. All these approaches rely on communications between neighbor nodes only, not to involve complex transmissions. However, this implies that different groups of neighbors must be connected to attain consensus over the whole network.

Distributed parameter estimation has been extensively studied for Wireless Sensors Networks (WSNs), where the low computational power of the nodes demands for simple operations and transmissions. Existing approaches can be classified as: incremental methods (Lopes & Sayed, 2007; Ram, Nedić, & Veeravalli, 2007; Sayed & Lopes, 2006); diffusion approaches (Arablouei, Doğaņay, Werner, & Huang, 2014; Cattivelli, Lopes, & Sayed, 2008; Cattivelli & Sayed, 2010; Nassif, Vlaski, & Sayed, 2019); *consensus+innovations* techniques (Kar & Moura, 2013; Sahu, Jakovetić, & Kar, 2018; Sahu, Kar, Moura, & Poor, 2016) and consensus-based distributed strategies, among which the approaches proposed in Mateos et al. (2009) and Schizas, Mateos, and Giannakis (2007, 2009) rely on the Alternating Direction Method of Multipliers (ADMM) (Boyd, Parikh, Chu, Peleato, & Eckstein, 2011). The main limitation of incremental methods lies in the need for a cycle in the network. Instead, the other approaches do not constrain the network topology, as they only require it to be connected. Nonetheless, most distributed techniques cannot handle constraints on the unknowns or scenarios in which the nodes have to reach consensus over a subset of the unknowns. These restrictions are overcome by the diffusion approach in Nassif et al. (2019) and the strategy in Sahu et al. (2018), respectively. *Consensus+innovations* approaches are further extended in Sahu et al. (2016) to handle the general framework in which the nodes observe a noisy nonlinear combination of the unknown parameters.

Even though distributed approaches have shown to perform well in different scenarios, they are generally not suited for applications in which the communication between devices is limited or impossible, *e.g.,* they cannot share information among each other for privacy reasons. On the other hand, with cloud technologies (Mell & Grance, 2011) becoming more widespread, many consumer devices have now on-demand access to repositories of shared information and common resources, with customizable computational power and memory, that can be accessed

☆ The material in this paper was partially presented at the 2018 American Control Conference, June 27–29, 2018, Milwaukee, WI, USA. This paper was recommended for publication in revised form by Associate Editor Luca Schenato under the direction of Editor Christos G. Cassandras.

* Corresponding author.
E-mail addresses: valentina.breschi@polimi.it (V. Breschi), alberto.bemporad@imtlucca.it (A. Bemporad), ilya@umich.edu (I.V. Kolmanovsky).

https://doi.org/10.1016/j.automatica.2020.109175
0005-1098/

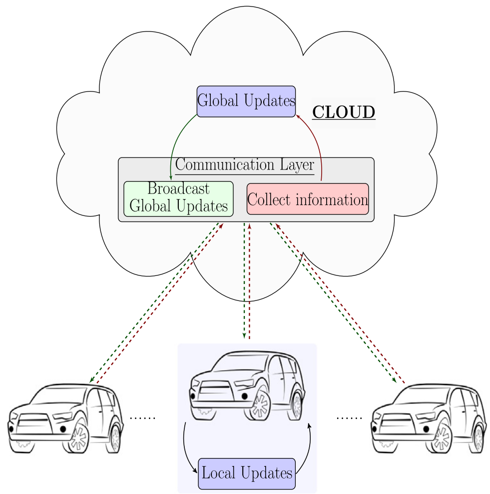


Fig. 1. Cloud-connected vehicles (Breschi, Kolmanovsky, & Bemporad, 2018).

and released with minimum effort. This motivates research into cloud-aided solutions. Indeed, cloud-based strategies have already been considered for automotive applications in Li et al. (2017, 2016) and Ozatay et al. (2014).

In a cloud aided setting, estimation can be carried out on board of each device, where data acquired locally are initially processed, and on the cloud, where the local estimates received from the nodes are refined to reach consensus on parameters common to all devices. As a motivating case study, consider a connected fleet of vehicles as in Fig. 1 and assume that a set of unknown parameters has to be estimated for the prognostics of automotive components such as fuel pumps (Taheri, Gusikhin, & Kolmanovsky, 2016) or brake pads (Howell et al., 2010). Under the assumptions that wear and fuel consumption models are known a priori and that the component wear-rate as a function of the workload[1] is common to similar devices, it is possible to use cloud-aided solutions to reach consensus on the wear-rates on the cloud, so that time (or mileage) to deplete remaining component life can be estimated and condition-based maintenance actions can be optimally scheduled.

Since we are interested in estimating the unknown parameters both locally and on the cloud, we present cloud-aided solutions for constrained collaborative least-squares estimation. Similarly to what is done in our conference paper (Breschi et al., 2018), we cast a separable optimization problem, that is solved via a dedicated instance of the Alternating Direction Method of Multipliers (ADMM). Nonetheless, differently from Breschi et al. (2018), we consider a broader range of scenarios, by devising solutions to address partial linear and nonlinear consensus constraints, and we provide theoretical guarantees for the case of linear consensus. Performance of all approaches is assessed with at least one numerical example, while a formal proof of convergence for the nonlinear scenario is left for future research. All these methods are tailored to retrieve both parameters common to all the nodes (*global*) and *local* parameters from data, while most approaches can only estimate common parameters. As in most existing solutions, we assume *ideal* transmissions between the devices and the cloud, namely we neither account for disturbances on the communication channels nor for losses of information.

The paper is organized as follows. The estimation problem is formalized in Section 2. The methods for linear and nonlinear collaborative estimation are then described in Sections 3–4. All the sections report the results of at least one simulation example, to show the performance of the methods. Concluding remarks and directions for future research are summarized in Section 5.

[1] Cumulative fuel flow or energy dissipated in the brakes

Notation. Let $\mathbb{N}$, $\mathbb{R}^+$ and $\mathbb{R}^n$ be the sets of natural numbers, real positive numbers (excluding zero) and real vectors of dimension n, respectively. Given a vector $a \in \mathbb{R}^n$, a_i indicates its ith element, $\|a\|_2$ is its Euclidean norm. The cardinality of a set $\mathcal{A}$ is indicated as $|\mathcal{A}|$, while $\mathcal{P}_{\mathcal{A}}(a)$ denotes Euclidean projection of the vector a onto $\mathcal{A}$. Let $A \in \mathbb{R}^{n\times m}$ be an $n \times m$ dimensional matrix, then A' denotes its transpose. The identity matrix of dimension n is indicated as I_n and the zero vector of dimension n is denoted as 0_n. Finally, $\mathcal{N}(\mu, \sigma^2)$ indicates a Gaussian distribution with mean μ and standard deviation σ.

2. Problem statement

Consider N devices (also referred to as *nodes* or *agents*) described by the same model, as it is reasonable to assume when considering mass produced devices. Suppose that all agents are connected to the cloud, with which they can exchange information. Furthermore, let the output $y_n(t) \in \mathbb{R}^{n_y}$ of the nth node at time t be given by

$$y_n(t) = X_n(t)'\theta_n^o + e_n(t), \quad n = 1, \dots, N, \tag{1}$$

where $\theta_n^o \in \mathbb{R}^{n_\theta}$ is an *unknown* parameter vector to be estimated from data, $X_n(t) \in \mathbb{R}^{n_\theta \times n_y}$ is the regressor (*e.g.,* a collection of past inputs and outputs of the nth system, if an autoregressive model is considered) and $e_n(t) \in \mathbb{R}^{n_y}$ is a zero-mean additive noise, independent of $X_n(t)$. We assume that the unknown parameters θ_n^o are *constant/slowly varying* and that they belong to *known* convex sets, *i.e.,* $\theta_n^o \in \mathcal{C}_n$, with $\mathcal{C}_n \subseteq \mathbb{R}^{n_\theta}$ for $n = 1, \dots, N$. As the N devices share the same model, we further suppose that there exists a set of unknown parameters $\theta^{g,o} \in \mathbb{R}^{n_g}$, with $n_g \leq n_\theta$, that is common to all agents.

Our goal is to estimate both the local parameters $\{\theta_n^o\}_{n=1}^N$ and the global parameter vector $\theta^{g,o}$ by exploiting the similarities between agents, the measurements available locally and the connection to the cloud. This estimation problem can be cast as follows:

$$\begin{aligned}
&\underset{\theta_1,\dots,\theta_N,\theta^g}{\text{minimize}} && \sum_{n=1}^{N} f_n(\theta_n) && \text{(2a)}\\
&\text{s.t.} && F(\theta_n) = \theta^g, \quad n = 1, \dots, N && \text{(2b)}\\
& && \theta_n \in \mathcal{C}_n, \quad n = 1, \dots, N && \text{(2c)}
\end{aligned}$$

where $F : \mathcal{F} \to \mathbb{R}^{n_g}$ is a *known* nonlinear mapping describing the relationship between the local and global parameters, which is assumed to be twice differentiable within its domain $\mathcal{F}$. The local fitting cost $f_n : \mathbb{R}^{n_\theta} \to \mathbb{R}$ is chosen as the least-squares function

$$f_n(\theta_n) = \frac{1}{2}\sum_{\tau=1}^{t} \left\| y_n(\tau) - X_n(\tau)'\theta_n \right\|_2^2, \quad n = 1, \dots, N. \tag{3}$$

Since we exploit the ADMM formalism (Boyd et al., 2011), problem (2) is equivalently recast as follows:

$$\begin{aligned}
&\underset{\substack{\theta_1,\dots,\theta_N\\ z_1,\dots,z_N,\theta^g}}{\text{minimize}} && \sum_{n=1}^{N} [f_n(\theta_n) + g_n(z_n)] && \text{(4a)}\\
&\text{s.t.} && F(\theta_n) = \theta^g, \quad n = 1, \dots, N, && \text{(4b)}\\
& && \theta_n = z_n, \quad n = 1, \dots, N, && \text{(4c)}
\end{aligned}$$

where $z_n \in \mathcal{C}_n$ is an auxiliary variable introduced to remove the parameter range constraint in (2) and g_n is the indicator function

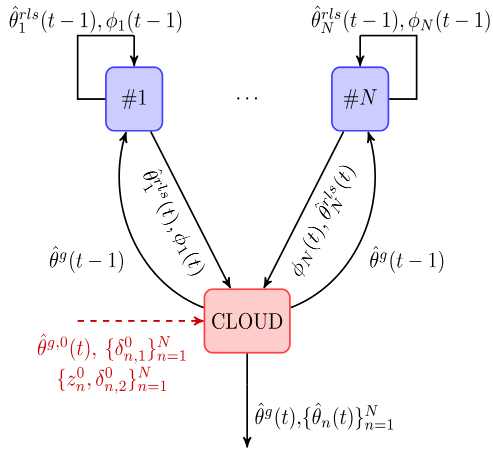


Fig. 2. Node-to-Cloud-to-Node transmissions at time t.

of $\mathcal{C}_n$, *i.e.*,

$$g_n(z_n) = \begin{cases} 0 \text{ if } z_n \in \mathcal{C}_n \\ +\infty \text{ otherwise}, \end{cases} \tag{5}$$

with $n = 1, \dots, N$. Even though no constraint is explicitly imposed on the value of θ^g in (4), a proper choice of $\{\mathcal{C}_n\}_{n=1}^N$ allows one to indirectly enforce the desired range constraints on the global parameters.

3. Linear consensus constraints

Let the consensus constraint in (4) be linear, i.e.,

$$F(\theta_n) = P\theta_n, \tag{6}$$

with $P \in \mathbb{R}^{n_g \times n_\theta}$ being a *known* matrix, that characterizes the relationship between the local and the global parameters. By appropriately defining the matrix P, problem (4) can be used to tackle different scenarios of practical interest, *e.g.*, for $P = I_{n_\theta}$ all unknowns are forced to take the same value.

As in our conference paper (Breschi et al., 2018), we solve problem (4), (6) via a recursive ADMM-based strategy. Specifically, at each time instant t, we solve the considered constrained optimization problem via a new instance of ADMM, while exploiting local estimates obtained at previous time steps with RLS. Differently from Breschi et al. (2018), we rely on the Node-to-Cloud-to-Node (N2C2N) communication scheme summarized in Fig. 2, thereby broadcasting the global estimates from the cloud to the nodes.

Since ADMM is directly implementable to solve the problem in (4) with the linear constraint (6), we can introduce the corresponding augmented Lagrangian:

$$\mathcal{L} = \sum_{n=1}^{N} \mathcal{L}_n(\theta_n, z_n, \delta_{n,1}, \delta_{n,2}, \theta^g), \tag{7a}$$

$$\begin{aligned} \mathcal{L}_n = {} & f_n(\theta_n) + g_n(z_n) + \delta_{n,1}'\varepsilon_{n,1} + \\ & + \delta_{n,2}'\varepsilon_{n,2} + \frac{\rho_1}{2}\|\varepsilon_{n,1}\|_2^2 + \frac{\rho_2}{2}\|\varepsilon_{n,2}\|_2^2, \end{aligned} \tag{7b}$$

with $\delta_{n,1}$, $\delta_{n,2}$ being the Lagrange multipliers associated with the constraints of the problem, $\rho_1, \rho_2 \in \mathbb{R}^+$ being tunable parameters and

$$\varepsilon_{n,1} = \theta_n - z_n, \quad \varepsilon_{n,2} = P\theta_n - \theta^g. \tag{8}$$

The ADMM steps to solve problem (4), (6) are the following:

$$\hat{\theta}_n^{k+1}(t) = \underset{\theta_n}{\operatorname{argmin}} \left\{ \mathcal{L}_n\left(\theta_n, z_n^k, \delta_{n,1}^k, \delta_{n,2}^k, \hat{\theta}^{g,k}(t)\right) \right\}, \tag{9a}$$

$$z_n^{k+1} = \mathcal{P}_{\mathcal{C}_n}\left(\hat{\theta}_n^{k+1}(t) + \frac{1}{\rho_1}\delta_{n,1}^k\right), \tag{9b}$$

$$\hat{\theta}^{g,k+1}(t) = \frac{1}{N}\sum_{n=1}^{N}\left(P\hat{\theta}_n^{k+1}(t) + \frac{1}{\rho_2}\delta_{n,2}^k\right), \tag{9c}$$

$$\delta_{n,1}^{k+1} = \delta_{n,1}^k + \rho_1(\hat{\theta}_n^{k+1}(t) - z_n^{k+1}), \tag{9d}$$

$$\delta_{n,2}^{k+1} = \delta_{n,2}^k + \rho_2(P\hat{\theta}_n^{k+1}(t) - \hat{\theta}^{g,k+1}(t)), \tag{9e}$$

where $k \in \mathbb{N}$ indicates the ADMM iteration. By examining the updates in (9b)–(9e), it is straightforward to observe that $\{z_n\}_{n=1}^N$, $\hat{\theta}^g(t)$, and $\{\delta_{n,1}, \delta_{n,2}\}_{n=1}^N$ can be updated iteratively at each time step t and ADMM iteration k. Moreover, thanks to the structure of the loss function f_n in (3), we can find the closed-form solution for problem (9a), so that the local estimate at time t can be computed as

$$\hat{\theta}_n^{k+1}(t) = \phi_n(t)\left\{\mathcal{Y}_n(t) + \xi_n^k\right\}, \tag{10}$$

with

$$\xi_n^k = \rho_1 z_n^k - \delta_{n,1}^k + P'(\rho_2\hat{\theta}^{g,k}(t) - \delta_{n,2}^k), \tag{11a}$$

$$\mathcal{Y}_n(t) = \sum_{\tau=1}^{t} X_n(\tau)y_n(\tau) = X_n(t)y_n(t) + \mathcal{Y}_n(t-1), \tag{11b}$$

$$\mathcal{X}_n(t) = \sum_{\tau=1}^{t} X_n(\tau)X_n(\tau)' = X_n(t)X_n(t)' + \mathcal{X}_n(t-1), \tag{11c}$$

$$\phi_n(t) = \left(\mathcal{X}_n(t) + \rho_1 I_{n_\theta} + \rho_2 P'P\right)^{-1}. \tag{11d}$$

The expression in (10) depends on both local information and quantities that are updated at each ADMM iteration k. To distinguish between these two terms, let $\hat{\theta}_n^{rls}(t)$ and $\hat{\theta}_n^{admm,k+1}(t)$ be given by

$$\hat{\theta}_n^{rls}(t) = \phi_n(t)\mathcal{Y}_n(t), \tag{12}$$

$$\hat{\theta}_n^{admm,k+1}(t) = \phi_n(t)\xi_n^k, \tag{13}$$

respectively, where only $\hat{\theta}_n^{admm,k+1}(t)$ changes at each ADMM iteration k, while only $\hat{\theta}_n^{rls}(t)$ relies explicitly on local measurements. By defining the partial estimate $\hat{\theta}_n^{rls}(t-1)$ as

$$\hat{\theta}_n^{rls}(t-1) = \phi_n(t-1)\mathcal{Y}_n(t-1), \tag{14}$$

with

$$\phi_n(t-1) = (\mathcal{X}_n(t-1) + \rho_1 I_{n_\theta} + \rho_2 P'P)^{-1},$$

straightforward manipulations lead to the following recursive formula:

$$\hat{\theta}_n^{rls}(t) = \hat{\theta}_n^{rls}(t-1) + K_n(t)\epsilon_n(t|t-1), \tag{15}$$

where

$$\epsilon_n(t|t-1) = y_n(t) - X_n(t)'\hat{\theta}_n^{rls}(t-1), \tag{16}$$

and the gain $K_n(t)$ is also recursively updated as

$$\mathcal{R}_n(t) = I_{n_\theta \times n_y} + X_n(t)'\phi_n(t-1)X_n(t), \tag{17a}$$

$$K_n(t) = \phi_n(t-1)X_n(t)(\mathcal{R}_n(t))^{-1}, \tag{17b}$$

$$\phi_n(t) = (I_{n_\theta} - K_n(t)X_n(t)')\phi_n(t-1). \tag{17c}$$

The iterative updates in (15)–(17) correspond to the ones of standard Recursive Least-Squares (RLS) (Ljung, 1999). It is thus straightforward to use the proposed approach, while relying on RLS estimators that could be available on board of each device already. Nonetheless, differently from standard RLS, the structure of matrix $\phi_n(t)$ in (11d) entails the presence of a regularization term in the minimized cost function, which is in turn shaped by the constraints of the problem. Note that these constraints are not explicitly enforced when computing $\hat{\theta}_n^{rls}(t)$. Since the formulas in (15)–(17) rely on the stream of local measurements only, it is reasonable to update $\hat{\theta}_n^{rls}(t)$ on board of the individual devices rather than on the cloud. However, $\hat{\theta}_n^{rls}(t)$ does not benefit from the estimate $\hat{\theta}^g(t-1)$ broadcast back from the cloud. To exploit this additional information, we introduce the *corrected* local estimate

$$\hat{\theta}_n^l(t) = \hat{\theta}_n^{rls}(t) + P'(\hat{\theta}^g(t-1) - P\hat{\theta}_n^{rls}(t-1)). \tag{18}$$

This estimate depends on information available at the node level and, thus, it can be computed by each device along with $\hat{\theta}_n^{rls}(t)$. At the same time, to avoid two-way transmissions between the nodes and the cloud at each ADMM iteration, the steps (9b)–(9e) are performed on the cloud, along with the computation of the refined estimate. The operations to be carried out at each time step t are outlined in Algorithm 1.

Remark 1. Algorithm 1 requires the initialization of the auxiliary variables and the global estimate. These initial values can be selected based on the local estimates $\{\hat{\theta}_n^{rls}(t)\}_{n=1}^N$. Indeed, at each instant t, z_n^0 can be chosen as the projection of $\hat{\theta}_n^{rls}(t)$ onto $\mathcal{C}_n$, while $\hat{\theta}^{g,0}(t)$ can be initialized as the sample mean of $\{P\hat{\theta}_n^{rls}(t)\}_{n=1}^N$. ■

Remark 2. Considering that a *perfect* synchronization between operations performed on the cloud and on board of all the devices is unachievable in practice, the outcome of Algorithm 1 could be influenced by delays due to this asynchronism. Nonetheless, if the actual parameters are constant or slowly varying, we expect that even asynchronous communications lead to sufficiently accurate estimates after some steps. ■

Algorithm 1 ADMM-RLS for partial consensus

Local inputs: Regressor/output pair $\{X_n(t), y_n(t)\}$; past estimates $\hat{\theta}_n^{rls}(t-1)$, $\phi_n(t-1)$, $\hat{\theta}^g(t-1)$; $\rho_1, \rho_2 \in \mathbb{R}^+$.
Cloud inputs: Initial values $\hat{\theta}^{g,0}(t)$, $\{\delta_{n,i}^0\}_{i=1}^2$ and z_n^0, with $n = 1, \dots, N$; $\rho_1, \rho_2 \in \mathbb{R}^+$.

Node-level computations

1. **each node** $n \in \{1, \dots, N\}$ **does**
   1.1. **update** $\phi_n(t)$ as in (17c);
   1.2. **update** $\hat{\theta}_n^{rls}(t)$ as in (15);
   1.3. **compute** $\hat{\theta}_n^l(t)$ as in (18);
   1.4. **transmit** $\hat{\theta}_n^{rls}(t)$ and $\phi_n(t)$ **to** the **cloud**;

Cloud-level computations

1. **iterate for** $k = 1, \dots$
   1.1. **for** $n = 1, \dots, N$
      1.1.1. **compute** $\hat{\theta}_n^{admm,k+1}(t)$ as in (13);
      1.1.2. **update** $\hat{\theta}_n^{k+1}(t)$ with (10);
      1.1.3. **compute** z_n^{k+1} as in (9b);
   1.2. **update** $\hat{\theta}^{g,k+1}(t)$ as in (9c);
   1.3. **for** $n = 1, \dots, N$
      1.3.1. **compute** $\delta_{n,1}^{k+1}$ as in (9d);
      1.3.2. **compute** $\delta_{n,2}^{k+1}$ as in (9e);
2. **until** the chosen stopping criterion is satisfied;
3. **transmit** $\hat{\theta}^g(t)$ **to** the **nodes**;

Local outputs: $\phi_n(t)$, local estimates $\hat{\theta}_n^{rls}(t)$ and $\hat{\theta}_n^l(t)$.
Cloud outputs: local estimates $\{\hat{\theta}_n(t)\}_{n=1}^N$; global estimate $\hat{\theta}^g(t)$.

3.1. Convergence analysis

As the approach summarized in Algorithm 1 is a particular instance of ADMM, we initially study the convergence rate and the asymptotic properties of Algorithm 1 at each time step t. In this case, we drop the dependence of the estimates on time to simplify the notation. Since the approach can also be seen as an extension of RLS tailored to exploit a cloud-aided framework, a bound on the local estimation error is derived and the consistency of local estimates is proven.

3.1.1. On the properties of Algorithm 1 at time t

Let $\mathcal{L}^o$ be the unaugmented Lagrangian associated to problem (4) with the constraints in (6), that depends on the measurements collected up to time t and is given by

$$\begin{aligned}\mathcal{L}^o &= \sum_{n=1}^N \mathcal{L}_n^o(\theta_n, z_n, \delta_{n,1}, \delta_{n,2}, \theta^g),\\ \mathcal{L}_n^o &= f_n(\theta_n) + g_n(z_n) + (\delta_{n,1})'\varepsilon_{n,1} + (\delta_{n,2})'\varepsilon_{n,2}.\end{aligned} \tag{19}$$

Assume that the following assumptions hold at each time instant t:

Assumption 1. There exists a saddle point for the unaugmented Lagrangian in (19), *i.e.*, there exists $\left(\{\theta_n^\star, z_n^\star, \delta_{n,1}^\star, \delta_{n,2}^\star\}_{n=1}^n, \theta^{g\star}\right)$ such that:

$$\mathcal{L}^o\left(\{\theta_n^\star, z_n^\star, \delta_{n,1}, \delta_{n,2}\}_1^N, \theta^{g\star}\right) \le \mathcal{L}^o\left(\{\theta_n^\star, z_n^\star, \delta_{n,1}^\star, \delta_{n,2}^\star\}_1^N, \theta^{g\star}\right)$$
$$\mathcal{L}^o\left(\{\theta_n^\star, z_n^\star, \delta_{n,1}^\star, \delta_{n,2}^\star\}_1^N, \theta^{g\star}\right) \le \mathcal{L}^o\left(\{\theta_n, z_n, \delta_{n,1}^\star, \delta_{n,2}^\star\}_1^N, \theta^{g}\right)$$

for all $\{\delta_{n,1}, \delta_{n,2}, \theta_n, z_n\}_{n=1}^N$ and for all θ^g. □

Assumption 2. Given finite initial values $\theta^{g,0}$ and $\{\delta_{n,1}^0, \delta_{n,2}^0, z_n^0\}_{n=1}^N$, the weighted sum

$$V^0 = \rho_2\|\hat{\theta}^{g,0} - \theta^{g\star}\|_2^2 + \frac{1}{N}\sum_{n=1}^N V_n^0, \tag{20a}$$

$$V_n^0 = \rho_1\|z_n^0 - z_n^\star\|_2^2 + \sum_{i=1}^2 \frac{1}{\rho_i}\|\delta_{n,i}^0 - \delta_{n,i}^\star\|_2^2, \tag{20b}$$

is finite. □

Remark 3. Since z_n^0, $\hat{\theta}^{g,0}$ and $\{\delta_{n,i}^0\}_{i=1}^2$ have to be chosen at every time instant, when a new instance of Algorithm 1 is carried out, Assumption 2 entails that the distance between the chosen initial conditions and the optimum is finite. This is likely to be verified for reasonably chosen initial conditions. As an example, if the initialization proposed in Remark 1 is used and the local estimates retrieved via RLS are converging to the true value, this assumption holds. ■

Inspired by Wei and Ozdaglar (2012), we derive a set of key inequalities that are verified at each ADMM iteration, established by the following lemmas.

Lemma 1. *Let* $\{\hat{\theta}_n^k, z_n^k, \delta_{n,1}^k, \delta_{n,2}^k\}_{n=1}^N$ *and* $\hat{\theta}^{g,k}$ *be generated by Algorithm 1, and introduce the differences*

$$r_{n,z}^{k+1} = z_n^{k+1} - z_n^k, \quad n = 1, \dots, N, \tag{21a}$$

$$r_g^{k+1} = \hat{\theta}^{g,k+1} - \hat{\theta}^{g,k}. \tag{21b}$$

Then, the following holds for all $k \in \mathbb{N}$ *and all* $\{\theta_n, z_n\}_{n=1}^N$ *and* θ^g*:*

$$\begin{aligned} &f_n(\hat{\theta}_n^{k+1}) + (\delta_{n,1}^{k+1} + \rho_1 r_{n,z}^{k+1})'\hat{\theta}_n^{k+1} + \\ &\quad + (\delta_{n,2}^{k+1} + \rho_2 r_g^{k+1})' P \hat{\theta}_n^{k+1} \leq f_n(\theta_n) + \\ &\quad + (\delta_{n,1}^{k+1} + \rho_1 r_{n,z}^{k+1})'\theta_n + (\delta_{n,2}^{k+1} + \rho_2 r_g^{k+1})' P \theta_n, \end{aligned} \tag{22a}$$

$$g_n(z_n^{k+1}) - (\delta_{n,1}^{k+1})' z_n^{k+1} \leq g_n(z_n) - (\delta_{n,1}^{k+1})' z_n, \tag{22b}$$

$$\left[\sum_{n=1}^N (\delta_{n,2}^{k+1})'\right] \hat{\theta}^{g,k+1} \geq \left[\sum_{n=1}^N (\delta_{n,2}^{k+1})'\right] \theta^g. \tag{22c}$$

Proof. Since $\hat{\theta}_n^{k+1}$ given by (9a) is the minimizer of $\mathcal{L}_n(\theta_n, z_n^k, \delta_{n,1}^k, \delta_{n,2}^k, \hat{\theta}^{g,k})$, it holds that

$$\begin{aligned} 0 \in\ &\partial f_n(\hat{\theta}_n^{k+1}) + \delta_{n,1}^k + P'\delta_{n,2}^k + \rho_1(\hat{\theta}_n^{k+1} - z_n^k) + \\ &+ \rho_2 P'(P\hat{\theta}_n^{k+1} - \hat{\theta}^{g,k}). \end{aligned}$$

By definition of subgradient, it is further verified that

$$f_n(\theta_n) \geq f_n(\hat{\theta}_n^{k+1}) + (\partial f_n(\hat{\theta}_n^{k+1}))'(\theta_n - \hat{\theta}_n^{k+1}),$$

for all θ_n. The proof of (22a) follows from algebraic manipulations of the previous inequalities, which involve the updates in (9d)–(9e). By exploiting a similar reasoning, it is straightforward to prove the relation in (22b). The inequality in (22c) can instead be derived from the optimality of $\hat{\theta}^{g,k+1}$. Then, by exploiting the update of the Lagrange multiplier in (9e), it can easily be seen that $\hat{\theta}^{g,k+1}$ is also defined as

$$\hat{\theta}^{g,k+1} = \underset{\theta^g}{\operatorname{argmin}} - \left[\sum_{n=1}^N (\delta_{n,2}^{k+1})'\right] \theta^g,$$

and the relation in (22c) easily follows. □

Lemma 2. *Let* $\{\hat{\theta}_n^k, z_n^k, \delta_{n,1}^k, \delta_{n,2}^k\}_{n=1}^N$ *and* $\hat{\theta}^{g,k}$ *be generated by Algorithm 1, and* p^k *and* p *be respectively defined as*

$$p^k = \sum_{n=1}^N \left[f_n(\hat{\theta}_n^k) + g_n(z_n^k)\right], \tag{23a}$$

$$p = \sum_{n=1}^N \left[f_n(\theta_n) + g_n(z_n)\right]. \tag{23b}$$

Then, for all $\{\theta_n, z_n\}_{n=1}^N$*, it holds that*

$$\begin{aligned} &p + p^{k+1} + \sum_{n=1}^N (\theta_n - \hat{\theta}_n^{k+1})' \left[\delta_{n,1}^{k+1} + P'\delta_{n,2}^{k+1} + \right. \\ &\left. + \rho_1 r_{n,z}^{k+1} + \rho_2 r_g^{k+1}\right] - \sum_{n=1}^N (z_n - z_n^{k+1})'\delta_{n,1}^{k+1} \geq 0, \end{aligned} \tag{24}$$

with $\{r_{n,z}^{k+1}\}_{n=1}^N$ *and* r_g^{k+1} *defined as in* (21a) and (21b), *respectively.*

Proof. The proof follows from straightforward manipulations of the relations (22a)–(22b). □

We can now prove that the convergence rate for Algorithm 1 at each time instant t is $O(\frac{1}{k})$.

Proposition 1. *Let* $\{\hat{\theta}_n^k, z_n^k, \delta_{n,1}^k, \delta_{n,2}^k\}_{n=1}^N$ *and* $\hat{\theta}^{g,k}$ *be generated by Algorithm 1, and* $\{\bar{\theta}_n^k, \bar{z}_n^k\}_{n=1}^N$ *and* $\bar{\theta}^{g,k}$ *be their ergodic averages up to the kth iteration,* i.e.,

$$\begin{aligned} &\bar{\theta}_n^k = \frac{1}{k}\sum_{s=0}^{k-1} \hat{\theta}_n^{s+1}, \quad \bar{z}_n^k = \frac{1}{k}\sum_{s=0}^{k-1} z_n^{s+1}, \quad n = 1, \dots, N, \\ &\qquad \bar{\theta}^{g,k} = \frac{1}{k}\sum_{s=0}^{k-1} \hat{\theta}^{g,s+1}. \end{aligned} \tag{25}$$

Under Assumptions 1–2, it holds that

$$0 \leq d^o \leq \frac{1}{k}\left(\frac{N}{2}V^0\right), \tag{26}$$

with V^0 *defined as in* (20a) *and* d^o *given by*

$$\begin{aligned} d^o = &\ \mathcal{L}^o\left(\{\bar{\theta}_n^k, \bar{z}_n^k, \delta_{n,1}^\star, \delta_{n,2}^\star\}_{n=1}^N, \bar{\theta}^{g,k}\right) + \\ &- \mathcal{L}^o\left(\{\theta_n^\star, z_n^\star, \delta_{n,1}^\star, \delta_{n,2}^\star\}_1^N, \theta^{g\star}\right). \end{aligned}$$

Proof. The left-hand-side of the relation in (26) directly results from the definition of saddle point. The rest of the proof can be found in Appendix A. □

By following the reasoning in Boyd et al. (2011), we can further state the following asymptotic result:

Proposition 2. *If Assumptions 1–2 hold, then*

(i) $\{\hat{\theta}_n^k, z_n^k, \delta_{n,1}^k, \delta_{n,2}^k\}_{n=1}^N$ *and* $\hat{\theta}^{g,k}$ *obtained by iterating the ADMM-RLS steps in* (9a)–(9e) *are bounded;*

(ii) residuals and objective convergence are attained, i.e.,

$$r_{n,1}^k = \hat{\theta}_n^k - z_n^k \underset{k\to\infty}{\longrightarrow} 0, \quad r_{n,2}^k = P\hat{\theta}_n^k - \hat{\theta}^{g,k} \underset{k\to\infty}{\longrightarrow} 0,$$

$$p^k \underset{k\to\infty}{\longrightarrow} \sum_{n=1}^N \left[f_n(\theta_n^\star) + g_n(z_n^\star)\right] = p^\star,$$

with p^k *defined as in* (23a). ■

Proof. The proof can be found in Appendix B. □

3.1.2. On the properties of the local estimates

We now study the properties of the local estimates $\hat{\theta}_n^k(t)$ only, by focusing on the characterization of the difference between $\hat{\theta}_n^k(t)$ and the actual value of the local parameters θ_n^o. To this end, we exploit the relations in (12)–(13), according to which

$$\hat{\theta}_n^{k+1}(t) = \hat{\theta}_n^{rls}(t) + \hat{\theta}_n^{admm,k+1}(t). \tag{27}$$

Due to its independence from the ADMM iterations, we initially consider $\hat{\theta}_n^{rls}(t)$ only, for which the following result holds.

Lemma 3. *Let* $V_n^{rls}(t)$ *be the Lyapunov-like function*

$$V_n^{rls}(t) = (\tilde{\theta}_n^{rls})' Q_n(t) \tilde{\theta}_n^{rls}, \tag{28}$$

with $\tilde{\theta}_n^{rls}(t) = \hat{\theta}_n^{rls}(t) - \theta_n^o$ *and* $Q_n(t) = (\phi_n(t))^{-1}$ *positive definite by definition. The following inequality holds:*

$$V_n^{rls}(t) \leq V_n^{rls}(0) + \sum_{\tau=1}^t (e_n(\tau))' e_n(\tau), \tag{29}$$

where $e_n(\tau)$ *is the additive white noise affecting the measurement at time* τ.

Proof. The proof follows from the result in Bittanti, Bolzern, and Campi (1990). Differently from Bittanti et al. (1990), we obtain the additional term $\sum_{\tau=1}^{t}(e_n(\tau))'e_n(\tau)$, that straightforwardly results from the introduction of noise in the definition of the output according to (1). □

Based on Lemma 3, we can now define a bound for the local estimates, that holds at each time instant t for a finite number of ADMM iterations k.

Proposition 3. *Let $\bar{\theta}_n^k(t)$, $\bar{z}_n^k(t)$ and $\bar{\theta}^{g,k}(t)$ be the ergodic averages in (25), obtained with k iterations of the new instance of Algorithm 1 carried out at time t. Denote with*

$$\bar{\delta}_{n,1}^k=\frac{1}{k}\sum_{s=0}^{k}\delta_{n,1}^{s+1}(t),\quad \bar{\delta}_{n,2}^k=\frac{1}{k}\sum_{s=0}^{k}\delta_{n,2}^{s+1}(t),$$

the averages of the Lagrange multipliers. Then, the following inequality holds:

$$\begin{aligned}\|\ \bar{\theta}_n^k(t)-\theta_n^\star\ \|_2^2\leq \frac{1}{\nu(t)}\Bigg\{&\alpha\|\tilde{\theta}_n^{rls}(0)\|_2^2+\sum_{\tau=1}^{t}(e_n(\tau))'e_n(\tau)+\\ &+\left[\eta_n(t)+\varphi_n^k(\hat{\theta}_n^{rls}(t),t)\right]^2+\beta(t)(\varphi_n^k(\theta_n^o,t))^2\Bigg\},\end{aligned}\tag{30}$$

where α and $\beta(t)$ are equal to the maximum eigenvalues of $Q_n(0)$ and $\phi_n(t)$, respectively, $\nu(t)$ is the smallest eigenvalue of $Q_n(t)$, and

$$\begin{aligned}\eta_n(t)=&\left\|\mathcal{Q}\phi_n(t)Q_n(0)\tilde{\theta}_n^{rls}(0)\right\|_2^2+\\ &+\left\|\mathcal{Q}\phi_n(t)\left(\sum_{\tau=1}^{t-1}Q_n(\tau)e_n(\tau)\right)\right\|_2^2+\|\mathcal{Q}K_n(t)e_n(t)\|_2^2,\end{aligned}$$

$$\begin{aligned}&\varphi_n^k(\theta,t)=\\ &=\psi_n^k(t)+\rho_1\left\|\theta-\frac{\bar{\delta}_{n,1}^k(t)}{\rho_1}\right\|_2^2+\rho_2\left\|P'\left(P\theta-\frac{\bar{\delta}_{n,2}^k(t)}{\rho_2}\right)\right\|_2^2+\frac{\omega_n^{0,k}(t)}{k},\end{aligned}$$

with $K_n(t)$ defined as in (17b), $\tilde{\theta}_n^{rls}(\tau)$ and $Q_n(\tau)$ introduced in Lemma 3, for $\tau=0,\dots,t$, and

$$\mathcal{Q}=(1-\rho_1)I_{n_\theta}-\rho_2P'P,\tag{31a}$$

$$\psi_n^k(t)=\rho_1\left\|\bar{z}_n^k(t)-\theta_n^o\right\|_2^2+\rho_2\left\|P'\left(\bar{\theta}^{g,k}(t)-P\theta_n^o\right)\right\|_2^2,\tag{31b}$$

$$\begin{aligned}\omega_n^{0,k}(t)=&\rho_1\left\|z_n^0(t)-z_n^k(t)\right\|_2^2+\rho_2\left\|P'\left(\hat{\theta}^{g,0}(t)-\hat{\theta}^{g,k}(t)\right)\right\|_2^2+\\ &+\left\|\delta_{n,1}^k(t)-\delta_{n,1}^0(t)\right\|_2^2+\left\|P'\left(\delta_{n,2}^k(t)-\delta_{n,2}^0(t)\right)\right\|_2^2.\end{aligned}\tag{31c}$$

Proof. The proof can be found in Appendix C. □

Remark 4. The upper-bound in (30) depends on the initial conditions $\hat{\theta}_n^{rls}(0)$ and $\phi_n(0)$ and the noise acting on the measurements. In addition, it depends on the performance of the ADMM over k runs, as it relies on the difference between the averages of the auxiliary and the global variable and their actual values (see (31b)). The upper-bound is also shaped by the deviations of the auxiliary variables, the global unknowns and the Lagrange multipliers over the k ADMM iterations performed at each time instant t (see (31c)). The influence of this last term is *inversely proportional* to the number of iterations. ■

From Proposition 2, at each time step and each ADMM iteration the estimates $\{\hat{\theta}_n^k, z_n^k, \delta_{n,1}^k, \delta_{n,2}^k\}_{n=1}^N$ and $\hat{\theta}^{g,k}$ are bounded. Given this result, we can also characterize the asymptotic properties of the local estimates $\hat{\theta}_n^k$. To this end, we introduce the following assumption.

Assumption 3. The regressor X_n and the noise e_n satisfy the following statistical properties:

(i) $\frac{1}{t}\mathcal{X}_n(t)\underset{t\to\infty}{\longrightarrow}\mathbb{E}[X_n(\tau)X_n(\tau)']=R_n\neq 0$,

(ii) $\sum_{\tau=1}^{t}X_n(\tau)e_n(\tau)\underset{t\to\infty}{\longrightarrow}\mathbb{E}[X_n(\tau)e_n(\tau)]=0$,

with $\mathcal{X}_n(t)$ defined as in (11c) and $\mathbb{E}[\cdot]$ denoting the expectation.

For an infinite estimation horizon, we can now prove the following consistency result:

Proposition 4. *Suppose Assumptions 1–3 hold and that the actual behavior of the each agent is described by (1). Then, $\hat{\theta}_n^k(t)$ is unbiased for $t\to\infty$.* ■

Proof. According to Eq. (10), $\hat{\theta}_n^k(t)$ is given by the combination of two terms. Consider $\hat{\theta}_n^{admm,k+1}(t)$, which is defined as in (13). Because of Assumption 3 and the boundedness of the auxiliary variables, global parameters and Lagrange multipliers (see Proposition 2), it follows that

$$\begin{aligned}&\hat{\theta}_n^{admm,k+1}(t)=\\ &\left(\frac{1}{t}\sum_{\tau=1}^{t}X_n(\tau)(X_n(\tau))'+\frac{\rho_1}{t}I+\frac{\rho_2}{t}PP'\right)^{-1}\frac{\xi_n^k}{t}\underset{t\to\infty}{\longrightarrow}0,\end{aligned}$$

for finite parameters ρ_1, ρ_2. We stress that this limit holds even if ADMM does not converge to the optimal value. Consider now the partial estimate $\hat{\theta}_n^{rls}(t)$ that is computed as in (15) with $K_n(t)=\phi_n(t)X_n(t)$. According to B. and J. (1999), this update is equal to the solution of the following optimization problem

$$\min_{\theta_n^{rls}}\frac{1}{2t}\left\{\sum_{\tau=1}^{t}\|y_n(\tau)-X_n(\tau)'\theta_n^{rls}\|_2^2+\Delta_n(0)\right\}.\tag{32}$$

where $\Delta_n(0)=(\theta_n^{rls}-\hat{\theta}_n^{rls}(0))\phi_n(0)^{-1}(\theta_n^{rls}-\hat{\theta}_n^{rls}(0))'$. The expression in (15) is thus equivalent to

$$\hat{\theta}_n^{rls}(t)=\left(\phi_n(0)^{-1}+\mathcal{X}_n(t)\right)^{-1}\left[\phi_n(0)^{-1}\hat{\theta}_n^{rls}(0)+\mathcal{Y}_n(t)\right],$$

with $\mathcal{Y}_n(t)$ given by (11b). If the initial parameters $\hat{\theta}_n^{rls}(0)$ and $\phi_n(0)$ are finite, the contribution of the initial condition vanishes for $t\to\infty$ and it is also verified that

$$\hat{\theta}_n^{rls}(t)\underset{t\to\infty}{\longrightarrow}R_n^{-1}\left[\frac{1}{t}\sum_{\tau=1}^{t}X_n(\tau)y_n(\tau)\right].\tag{33}$$

By replacing $y_n(t)$ with Eq. (1) and accounting again for Assumption 3, consistency easily follows. □

3.2. Example 1: full consensus

Consider a simple estimation scenario in which $N=5$ agents connected to the cloud are available to estimate the parameter vector $\theta\in\mathbb{R}^5$, with $\theta=[\,20\ 5\ 10\ 15\ 8\,]'$. Each agent observes a noisy combination of the unknowns, namely

$$y_n(t)=H_n'\theta+e_n(t),$$

with $H_n\in\mathbb{R}^5$ randomly generated according to the uniform distribution and $e_n\sim\mathcal{N}(0,25)$. The effect of noise on the measurements is assessed via the Signal-to-Noise Ratio (SNR), *i.e.*,

$$\text{SNR}_n=10\log\frac{\sum_{t=1}^{T}(y_n(t)-e_n(t))^2}{\sum_{t=1}^{T}e_n(t)^2}\ \text{dB},\ \ n=1,\dots,N,\tag{34}$$

which is around 15 dB for all agents. The unknowns are estimated over a horizon $T=1000$ without imposing any range constraint. The parameters used in Algorithm 1 are indicated in Table 1 and the remaining initial values are set according to Remark 1. Note that $P=I_5$, since the whole parameter vector is assumed to be shared by all agents. At the cloud level, the

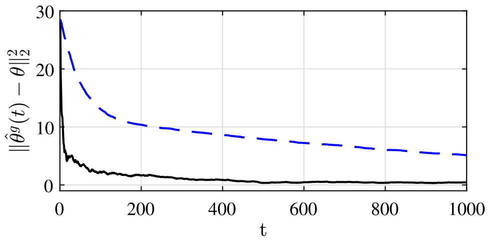


Fig. 3. Example 1: norm of the global estimation error. ADMM-RLS (black) and CLMS (Cattivelli & Sayed, 2010) (dashed blue).

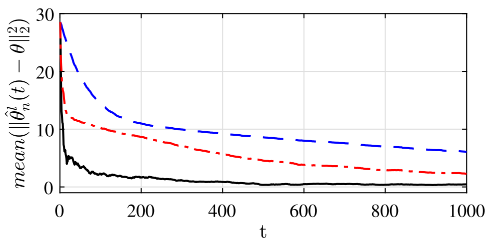


Fig. 4. Example 1: average norm of the local estimation error. ADMM-RLS (black), DLMS (Cattivelli & Sayed, 2010) (dashed blue) and *consensus+innovations* (Kar & Moura, 2013) (dotted–dashed red).

Table 1
Example 1: ADMM-RLS parameters.

$\hat{\theta}_n^{rls}(0)$	$\phi_n(0)$	ρ_2	$\delta_{n,1}^0$	$\delta_{n,2}^0$
0_5	$0.1I_5$	1	$10^{-3}I_5$	$10^{-3}I_5$

Table 2
Example 2: ADMM-RLS parameters.

$\phi_n(0)$	ρ_1	ρ_2	$\delta_{n,1}^0$	$\delta_{n,2}^0$
$0.1I_3$	1	0.1	$10^{-3}I_3$	$10^{-3}I_2$

performance of ADMM-RLS is assessed via the global estimation error, whose Euclidean norm over time is reported in Fig. 3. Our results are compared with the ones of the centralized least-mean squares (CLMS) (Cattivelli & Sayed, 2010). Fig. 3 clearly shows that ADMM-RLS outperforms CLMS in terms of accuracy. Let us further assume that all agents are also connected to each others. The performance attained at the node level with ADMM-RLS is thus compared with the one achieved with two distributed approaches, namely diffusion least-mean squares (DLMS) (Cattivelli & Sayed, 2010) and the *consensus+innovations* estimator introduced in Kar and Moura (2013),[2] by looking at average Euclidean norm of the local estimation error shown in Fig. 4. It can be observed that our approach performs better than the considered distributed estimation methods. For the comparison to be fair, the local estimation error is computed with respect to the corrected local estimate defined in (18), while, when using the *consensus+innovations* approach, the noise covariance is estimated as explained in Sahu et al. (2018).

3.3. Example 2: Partial consensus

Consider now $N = 100$ dynamical systems described by the Auto-Regressive model with eXogenous inputs (ARX)

$$y_n(t) = \theta_1^g y_n(t-1) + \theta_{n,2} y_n(t-2) + \theta_2^g u_n(t-1) + e_n(t).$$

[2] In running CLMS, DLMS and the *consensus+innovations* approach we start from the same initial conditions indicated in Table 1.

Table 3
Example 2: average RMSEs on the local estimates.

Mean(RMSE$_{\hat{\theta}_n^{rls}}$)	Mean(RMSE$_{\hat{\theta}_n^{l}}$)	Mean(RMSE$_{\hat{\theta}_n}$)
$2.3 \cdot 10^{-1}$	$0.9 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$

The parameter vector $\theta^g = [\,0.2\ 0.8\,]'$ is common to all the agents, while $\{\theta_{n,2}\}_{n=1}^N$ are purely local parameters sampled from the distribution $\mathcal{N}(0.4, 0.05^2)$. The exogenous inputs u_n are generated as sequences of i.i.d. elements uniformly distributed in the interval $[2,\ 3]$, while $e_n \sim \mathcal{N}(0, R_n)$ are white noise sequences with random covariance. We consider $R_n \in [1,\ 16]$, for $n = 1, \dots, N$, so that $\text{SNR}_n \in [6.4,\ 16.1]$ dB. A similar example is considered in our conference paper (Breschi et al., 2018). ADMM-RLS is used to iteratively estimate both local and global parameters over a horizon $T = 5000$ from input/output samples, under the assumptions that the values of the local parameters are known with an uncertainty of ± 0.1 and that the following priors are available on the global unknowns:

$$0.19 < \theta_1^g < 0.21, \quad 0.79 < \theta_2^g < 0.81. \tag{35}$$

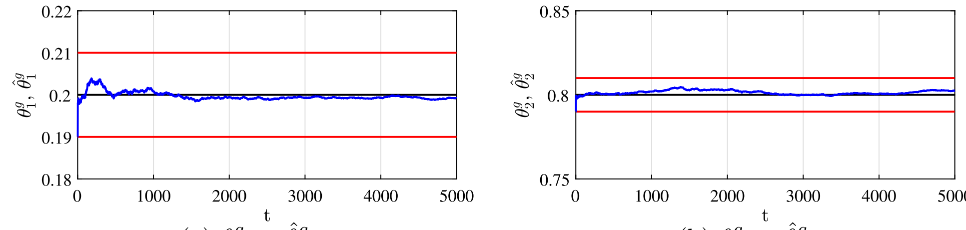
Algorithm 1 is run with the parameters in Table 2 and the initial local estimates $\hat{\theta}_n^{rls}(0)$ randomly chosen in the interval $[-10,\ 10]$. The initial values for the global estimates and the auxiliary variables are selected as explained in Remark 1. The global estimates retrieved on the cloud are reported in Fig. 5, showing that the global unknowns are accurately estimated and that they overall satisfy the range constraints, even though they are not directly enforced on the global estimates. The accuracy of the local estimates is assessed via the Root Mean Square Error (RMSE), defined as

$$\text{RMSE}_{\hat{\theta},i} = \sqrt{\frac{\sum_{t=1}^T \left(\theta_i - \hat{\theta}_i(t)\right)^2}{T}}, \tag{36}$$

where θ_i and $\hat{\theta}_i$ indicate the ith components of a generic parameter vector θ and its estimate, respectively. Table 3 reports the average RMSEs on the local estimates. At the node level, it is clear that $\hat{\theta}_n^l$ is more accurate than $\hat{\theta}_n^{rls}$, highlighting the benefit of using the proposed communication scheme. At the same time, $\hat{\theta}_n^{rls}$ is less accurate than the one refined on the cloud, since $\hat{\theta}_n$ relies on information gathered from all the agents. On the contrary, $\hat{\theta}_n^l$ is slightly more accurate than $\hat{\theta}_n$. This result is somehow expected, since $\hat{\theta}_n^l$ exploits the global estimate computed on the cloud.

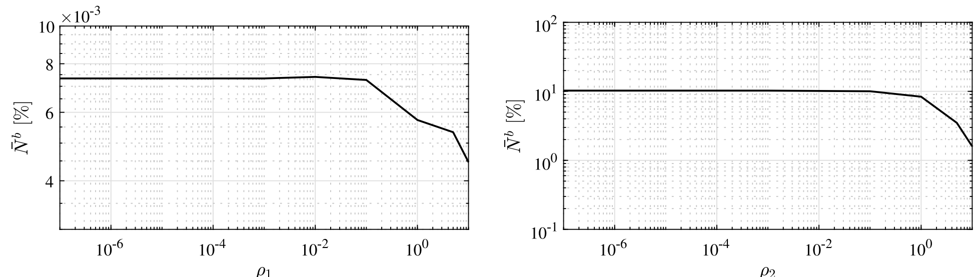
3.3.1. Example 2: Sensitivity analysis

The chosen penalty parameters ρ_1 and ρ_2 directly impact the performance of ADMM. Therefore, we perform a posterior sensitivity analysis, by alternatively changing the value of one of the tuning parameters, while fixing the other to its value in Table 2. We evaluate the influence of these tuning parameters on the satisfaction of the constraints via two performance indexes. On one hand, we consider the average percentage of parameter range constraint violations $\bar{N}^b$, which is computed by considering only cases when the estimate exceeds the range by more than $\pm 10^{-4}$. On the other hand, the quality of the global estimate is assessed the Euclidean of RMSE$_{\hat{\theta}^g}$. Figs. 6–7 show how these performance indexes for different choices of ρ_1 and ρ_2, respectively. Since ρ_1 directly influences the enforcement of the range constraints, it is clear that the higher its value is, the lower is the percentage of constraint violations. Instead, ρ_2 positively affects violation constraints for choices of $\rho_2 \in [10^{-1},\ 10]$. On the other hand, the value of $\|\text{RMSE}_{\hat{\theta}^g}\|_2$ tends to decreases for $\rho_1 \in [10^{-3},\ 5]$, while the accuracy of the global estimate visibly deteriorate for $\rho_2 > 5$. These results indicate the need for a good trade-off between the weights associated to the different constraints, while particular care has to be taken in selecting ρ_2.



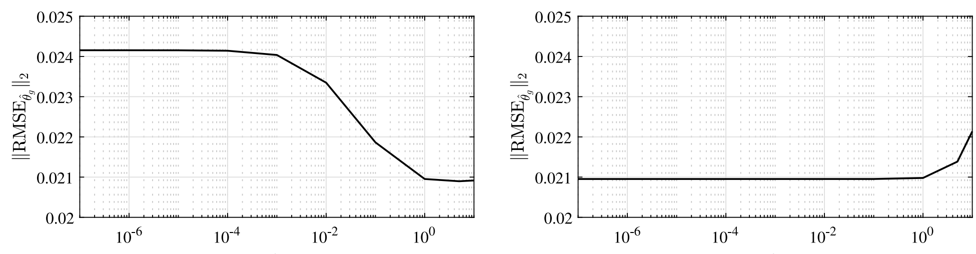
(a) θ_1^g vs $\hat{\theta}_1^g$ (b) θ_2^g vs $\hat{\theta}_2^g$

Fig. 5. Example 2: true vs estimated global parameters. True (black), ADMM-RLS (blue) and bounds (red, see (35)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) $\bar{N}^b$ [%] vs ρ_1, with $\rho_2 = 0.1$ (b) $\bar{N}^b$ [%] vs ρ_2, with $\rho_1 = 1$

Fig. 6. Example 2: average percentage of range constraints violations $\bar{N}^b$ [%] vs ρ_1 and ρ_2.



(a) $\|\text{RMSE}_{\hat{\theta}^g}\|_2$ vs ρ_1, with $\rho_2 = 0.1$ (b) $\|\text{RMSE}_{\hat{\theta}^g}\|_2$ vs ρ_2, with $\rho_1 = 1$

Fig. 7. Example 2: 2-norm of the RMSE, $\|\text{RMSE}^g\|_2$, vs ρ_1 and ρ_2.

4. Nonlinear consensus constraints

Suppose now that the *known* relationship between the global and local parameters is nonlinear, as in problem (2). For this more involved case, we still aim at devising a recursive method to estimate the unknown parameters, while exploiting the communication scheme in Fig. 2. Similarly to the linear case in Section 3, at each time instant we exploit ADMM. Even though ADMM can be directly applied to the problem, due to the nonlinear consensus constraint the resulting scheme requires the solution of a nonlinear problem to update the local estimate at each time step and ADMM iteration. This might be impractical, especially when the nonlinear problem has to be solved locally. A possible solution to simplify the iterations is to adopt an approximate ADMM scheme, *e.g.*, the approach proposed in Benning, Knoll, Schönlieb, and Valkonen (2015). When applied to our problem, the approach in Benning et al. (2015) relies on the substitution of $F(\theta_n)$ with its Taylor expansion around $\hat{\theta}_n^k$. Although this choice allows us to simplify the operations performed locally, the resulting scheme still has some undesirable features. In particular, it entails that all the estimates have to be updated at each ADMM iteration, requiring a prohibitive number of back and forth communications. To handle this additional problem, we exploit the constraint in (4c), and we modify the consensus condition as

$$F(z_n) = \theta^g, \quad \forall n \in \{1, \ldots, N\}. \tag{37}$$

This leads to the following ADMM-based scheme:

$$\hat{\theta}_n^{k+1}(t) = \underset{\theta_n}{\text{argmin}} \left\{ f_n(\theta_n) + \delta_{n,1}^{k'}(\theta_n - z_n^k) + \frac{\rho_1}{2}\|\theta_n - z_n^k\|_2^2 \right\}, \tag{38a}$$

$$z_n^{k+1} = \underset{z_n}{\text{argmin}} \left\{ g_n(z_n) + \delta_{n,1}^{k'}\varepsilon_{n,1}^{k+1} + \frac{\rho_1}{2}\|\varepsilon_{n,1}^{k+1}\|_2^2 + \right. \\ \left. + \delta_{n,2}^{k'}\varepsilon_{n,2}^{k} + \frac{\rho_2}{2}\|\varepsilon_{n,2}^{k}\|_2^2 \right\}, \tag{38b}$$

$$\hat{\theta}^{g,k+1}(t) = \frac{1}{N}\sum_{n=1}^{N}\left[F(z_n^{k+1}) + \frac{1}{\rho_2}\delta_{n,2}^k\right], \tag{38c}$$

$$\delta_{n,1}^{k+1} = \delta_{n,1}^k + \rho_1(\hat{\theta}_n^{k+1}(t) - z_n^{k+1}), \tag{38d}$$

$$\delta_{n,2}^{k+1} = \delta_{n,2}^k + \rho_2(F(\hat{\theta}_n^{k+1}(t)) - \hat{\theta}^{g,k+1}(t)), \tag{38e}$$

with $\varepsilon_{n,1}^{k+1} = \hat{\theta}_n^{k+1}(t) - z_n$ and $\varepsilon_{n,2}^k = F(z_n) - \hat{\theta}^{g,k}$. The closed form solution for (38a) is given by (10), *i.e.*,

$$\hat{\theta}_n^{k+1}(t) = \phi_n(t)\left\{\mathcal{Y}_n(t) + \xi_n^k\right\},$$

with $\xi_n^k = \rho_1 z_n^k - \delta_{n,1}^k$, $\phi_n(t) = \left(\mathcal{X}_n(t) + \rho_1 I_{n_\theta}\right)^{-1}$ and $\mathcal{Y}_n(t)$ and $\mathcal{X}_n(t)$ defined as in (11b) and (11c), respectively. As in Section 3, we can introduce the partial estimates $\hat{\theta}_n^{rls}(t)$ and $\hat{\theta}_n^{admm,k}(t)$, given by (12) and (13), respectively. Therefore, it is easy to prove that $\hat{\theta}_n^{rls}$ can be iteratively updated as in (15)–(17). The manipulation of the consensus constraint thus leads to simple iterative updates for the local estimates, which are independent of the nonlinear function F. As in the linear case, the local estimate $\hat{\theta}_n^{rls}(t)$ is still independent of estimate broadcast from the cloud to the nodes. Therefore, also in this setting we introduce the *corrected* local estimate $\hat{\theta}_n^{l}(t)$, given by

$$\hat{\theta}_n^{l}(t) = \hat{\theta}_n^{rls}(t) + G(\hat{\theta}^g(t-1) - F(\hat{\theta}_n^{rls}(t-1)), \hat{\theta}_n^{rls}(t-1)). \tag{39}$$

The function $G : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_g} \to \mathbb{R}^{n_\theta}$ returns a vector of the same dimension as $\hat{\theta}_n^{rls}$, with entries equal to 0 in correspondence of purely local parameters, while the other components are properly shaped by the difference between the global estimate and purely local approximations of the global parameters.

Consider now the update of the auxiliary variable in (38b), which depends on F. Instead of solving the nonlinear optimization problem in (38b), we linearize $F(z_n)$ around z_n^k, *i.e.*,

$$F(z_n) \approx F(z_n^k) + A_n^{k+1}(z_n - z_n^k), \tag{40}$$

with

$$A_n^{k+1} = \left.\frac{\partial F(z_n)}{\partial z_n}\right|_{z_n = z_n^k}. \tag{41}$$

The resulting function to be minimized is

$$\hat{\mathcal{L}}_n = g_n(z_n) + \delta_{n,1}^{k'}(\hat{\theta}_n^{k+1}(t) - z_n) + \frac{\rho_1}{2}\|\hat{\theta}_n^{k+1}(t) - z_n\|_2^2 + \\ + \delta_{n,2}^{k'}(A_n^{k+1}z_n - c_n^{k+1}) + \frac{\rho_2}{2}\|A_n^{k+1}z_n - c_n^{k+1}\|_2^2,$$

with $c_n^{k+1} = A_n^{k+1}z_n^k + \hat{\theta}^{g,k}(t) - F(z_n^k)$. It is straightforward to prove that the auxiliary variables can be updated as follows

$$\begin{aligned} z_n^{k+1} &= \mathcal{P}_{\mathcal{C}_n}\left(Z_1^{-1}Z_2\right), \\ Z_1 &= \rho_1 I_{n_\theta} + \rho_2 A_n^{k+1'} A_n^{k+1}, \\ Z_2 &= \delta_{n,1}^k + \rho_1\hat{\theta}_n^{k+1}(t) + A_n^{k+1'}\left(\rho_2 c_n^{k+1} - \delta_{n,2}^k\right). \end{aligned} \tag{42}$$

An outline of an iteration of the approximated ADMM scheme is provided in Algorithm 2. Note that no inner loop is introduced for the solution of a nonlinear optimization problem.

Remark 5. By adding preconditioning within our ADMM-based approach, the proposed method corresponds to the one in Benning et al. (2015). Therefore, we expect that the properties of Algorithm 2 can be characterized with the same reasoning as in Valkonen (2014). We stress that the proofs in Valkonen (2014) rely on some technical assumptions, which are satisfied in our setting. Indeed, F is assumed to be twice differentiable in its domain and the cost function is convex, proper and lower semi-continuous. We remark that a new instance of Algorithm 2 is initialized and carried out at each time instant t, so that the arguments in Valkonen (2014) allow us to characterize the performance of the approach at a given time instant only. To delineate the behavior of Algorithm 2 over time, an upper-bound on the local error can be derived. This bound is expected to resemble the one in (30), since local estimates are decomposed in a similar way. Nonetheless, the bound is likely to be shaped by the different nature of the consensus constraint and, in particular, by the error performed in locally linearizing the constraint. ■

Algorithm 2 ADMM-RLS for nonlinear consensus

Local inputs: Regressor/output pairs $\{X_n(t), y_n(t)\}$, past estimates $\hat{\theta}_n^{rls}(t-1)$, $\phi_n(t-1)$, $\hat{\theta}^g(t-1)$; $\rho_1, \rho_2 \in \mathbb{R}^+$.
Cloud inputs: Initial values $\hat{\theta}^{g,0}(t)$, $\{\delta_{n,i}^0\}_{i=1}^2$ and z_n^0, $n = 1, \ldots, N$; $\rho_1, \rho_2 \in \mathbb{R}^+$.

Node-level computations

1. **each node** $n \in \{1, \ldots, N\}$ **does**
   - 1.1. **update** $\phi_n(t)$ as in (17c);
   - 1.2. **update** $\hat{\theta}_n^{rls}(t)$ as in (15);
   - 1.3. **compute** $\hat{\theta}_n^{l}(t)$ as in (39);
   - 1.4. **transmit** $\hat{\theta}_n^{rls}(t)$ and $\phi_n(t)$ **to** the **cloud**;

Cloud-level computations

1. **iterate for** $k = 1, \ldots$
   - 1.1. **for** $n = 1, \ldots, N$
     - 1.1.1. **compute** $\hat{\theta}_n^{admm,k+1}(t)$ as in (13);
     - 1.1.2. **update** $\hat{\theta}_n^{k+1}(t)$ with (10);
     - 1.1.3. **compute** A_n^{k+1} as in (41);
     - 1.1.4. **compute** z_n^{k+1} as in (42);
   - 1.2. **update** $\hat{\theta}^{g,k+1}(t)$ as in (38c);
   - 1.3. **for** $n = 1, \ldots, N$
     - 1.3.1. **compute** $\delta_{n,1}^{k+1}$ as in (38d);
     - 1.3.2. **compute** $\delta_{n,2}^{k+1}$ as in (38e);
2. **until** the chosen stopping criterion is satisfied;
3. **transmit** $\hat{\theta}^g(t)$ **to** the **nodes**;

Local outputs: $\phi_n(t)$, local estimates $\hat{\theta}_n^{rls}(t)$ and $\hat{\theta}_n^{l}(t)$.
Cloud outputs: local estimates $\{\hat{\theta}_n(t)\}_{n=1}^N$; global estimate $\hat{\theta}^g(t)$.

4.1. Example 3: estimation over a fleet of vehicles

An accurate estimate of the mass, road grade and drag coefficient of a vehicle is of paramount importance in many vehicle control applications (Vahidi & Eskandarian, 2003). Therefore, methods for effective estimation of these unknowns have been extensively studied within the control community (Kidambi, Harne, Fujii, Pietron, & Wang, 2014; Vahidi, Stefanopoulou, & Peng, 2005). Nonetheless, estimation is generally carried out by using information collected from one vehicle only, without exploiting the similarities between vehicles. In this example we do exploit this additional information, by considering the problem of mass, road grade and drag coefficient estimation over a fleet of N vehicles. In our simulations, we assume that the velocities of the N vehicles are measured locally and that at least an estimate of the longitudinal force acting on each vehicle is available. The considered unknowns can thus be estimated by relying on the model of the vehicle longitudinal dynamics (Rajamani, 2012). Accordingly, by considering the ideal setting in which the unknowns are constant over the estimation horizon and the wind velocity is negligible, the regressor/output relationship for the nth vehicle is described by the following ARX model:

$$y_n(t) = X_n(t)'\theta(t) + e_n(t),$$

with output $y_n(t) = v_{x,n}(t) - v_{x,n}(t-1) + T_s fg$, where $v_{x,n}$ [m/s] is the vehicle velocity and $T_s = 0.1$ s is the sampling time. The parameter vector $\theta_n \in \mathbb{R}^3$ is given by

$$\theta_n = \begin{bmatrix} \frac{1}{m_n} & \frac{C_d}{m_n} & \sin(\beta_n) \end{bmatrix}',$$

with m_n [kg], C_d and β_n [rad] being the mass, drag coefficient and road grade for the nth vehicle, respectively. The regressor $X_n(t)$ is

Table 4
Example 3: Parameters of the nth vehicle.

Sym	Meaning and UoM	Value
m_n	Mass [kg]	$\mu_n \sim \mathcal{N}(1500, 10^4)$
β_n	Road grade [rad]	$\beta_n \sim \mathcal{N}(0, 8 \cdot 10^{-5})$
C_d	Drag coefficient	0.4
ϱ	Air density [kg/m^3]	1.18
A_f	Frontal area [m^2]	3
f	Rolling resistant coefficient	0.015
g	Gravitational acceleration [m/s^2]	9.81

Table 5
Example 3: ADMM-RLS parameters.

$\hat{\theta}_n^{rls}(0)$	$\phi_n(0)$	ρ_1	ρ_2	$\delta_{n,1}^0$	$\delta_{n,2}^0$
$\begin{bmatrix} 10^{-2} \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 10^{-4} & 0 & 0 \\ 0 & 10^{-5} & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix}$	10^{-2}	10^{-3}	$10^{-3}I_5$	$10^{-3}I_5$

equal to

$$X_n(t) = T_s \left[F_{x,n}(t-1) \quad -0.5\varrho A_f v_{x,n}(t-1)^2 \quad -g \right]',$$

where $F_{x,n}$ [N] is the longitudinal force acting on the nth vehicle. The remaining parameters are reported in Table 4. The additive sequence $e_n(t)$ is introduced to account for possible model mismatch and for the noise acting on the measurements.[3] In practice, it is reasonable to assume masses and road grades to be different for each vehicle, so that they can be treated as purely local parameters. At the same time, given the similarities between the N vehicles in the fleet, the drag coefficient can be reasonably assumed to be equal for all vehicles, thus representing a global parameter. The nonlinear consensus constraint is enforced by imposing the following nonlinear relationship between the local parameters:

$$F(\theta_n) = (\theta_{n,1})^{-1}\theta_{n,2}, \tag{43}$$

while the function G in (39) returns the three dimensional vector $\left[0 \; \hat{\theta}_{n,1}^{rls}(t-1)\hat{\theta}^g(t-1) - \hat{\theta}_{n,2}^{rls}(t-1) \; 0 \right]'$. According to the physical meaning of the estimated parameters and to guarantee that A_n^{k+1} in (42) is well defined, the local parameters are constrained in the sets $\mathcal{C}_n \subset \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}$, for $n = 1, \dots, N$. We acquire measurements from a minimum of $N_{min} = 1$ to a maximum of $N_{max} = 20$ vehicles, so that the performance of the proposed method can be assessed for fleets of different sizes. The available velocity profiles resemble the ones reported in Fig. 8, indicating that there can be zero-velocity intervals, especially at the beginning of the estimation horizon. Algorithm 2 is used to estimate the unknowns over $T = 670$ s, with the parameters reported in Table 5. The auxiliary variables and global estimates initialized following the same reasoning in Remark 1. Fig. 9 shows the Root Mean Square Error (RMSE) (36) for the global estimate. This result indicates that accuracy generally improves when N increases. However, the evolution of the RMSE over N is non-monotone due to the presence of *critical* agents, *i.e.,* vehicles with considerable zero-velocity periods. Additional insights are given by the absolute estimation errors on the global parameter, reported in Fig. 10. It is clear that the estimate obtained with a fleet of $N = 20$ vehicle is more accurate than the ones obtained for $N = 1, 2$, with a visible improvement in the quality of the estimated drag coefficient when at least 2 vehicles are connected to the cloud.

[3] In simulating the longitudinal behavior of each vehicle, we have corrupted that the system through a process noise $w \sim \mathcal{N}(0, 0.01)$, while the measurements are altered by a Gaussian zero-mean white process with standard deviation 0.01.

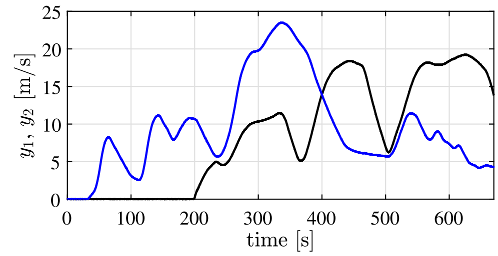


Fig. 8. Example 3: sample velocity profiles. Black: y_1, blue: y_2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

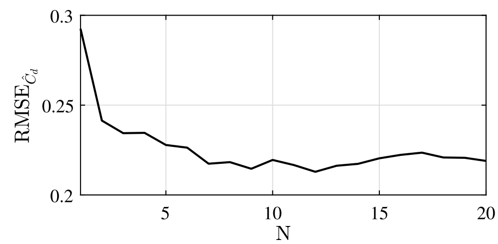


Fig. 9. Example 3: RMSE$_{\hat{C}_d}$ vs dimension of the fleet N.

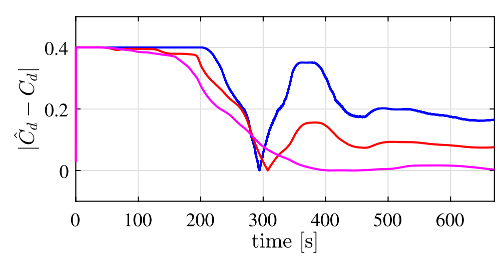


Fig. 10. Example 3: absolute estimation error $|\hat{C}_d - C_d|$. Blue: $N = 1$, red: $N = 2$, magenta: $N = 20$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Example 3: average RMSEs on the mass and the road grade.

Mean(RMSE$_{\hat{m}_n}$) [kg]	Mean(RMSE$_{\hat{\beta}_n}$) [rad]
446.9	$7.6 \cdot 10^{-3}$

Note that, for $N = 1$, we use standard RLS (Ljung, 1999), without imposing range constraints on the parameters. The average RMSEs on the mass and road grade are reported in Table 6, so to assess the performance of the approach at the node level. It is worth pointing out that the initial error on the value of the mass is around 1400 [kg].

5. Conclusion

We presented two ADMM-based methods for recursive constrained collaborative estimation. These approaches allow us to handle linear and nonlinear consensus constraints, with only a set of (simple) computations executed locally. More complex tasks are instead performed by a centralized resource on the cloud. Future investigations will be devoted to a theoretical analysis of the method proposed for the nonlinear case. On the methodological side, future research will address developing solutions to consider more general consensus constraint. Due to the demonstrated sensitivity of the method to the tuning parameters in the augmented Lagrangian, auto-tuning strategies will also be investigated.

Appendix A. Proof of Proposition 1

Since the result of Lemma 2 holds for all $\{\theta_n, z_n\}_{n=1}^N$, θ^g and for all ADMM iterations k, it follows that

$$p^\star - p^{s+1} + \rho_2\left[\sum_{n=1}^N (d_{\theta_n}^{s+1})'\right]P' r_g^{s+1} - \sum_{n=1}^N (d_{z_n}^{s+1})'\delta_{n,1}^{s+1} + \sum_{n=1}^N (d_{\theta_n}^{s+1})'\left(\delta_{n,1}^{s+1} + P'\delta_{n,2}^{s+1} + \rho_1 r_{n,z}^{s+1}\right) \geq 0, \tag{A.1}$$

where

$$d_{\theta_n}^{s+1} = \hat{\theta}_n^\star - \hat{\theta}_n^{s+1}, \quad d_{z_n}^{s+1} = z_n^\star - z_n^{s+1}, \quad n = 1, \ldots, N, \tag{A.2}$$

and $r_{n,z}^{s+1}$ and r_g^{s+1} are defined as in (21a) and (21b), respectively. By adding and subtracting $(\theta^{g\star})'\left[\sum_{n=1}^N \delta_{n,2}^{s+1}\right]$ to (A.1) and by exploiting the relation in (22c) and the primal feasibility conditions, simple algebraic manipulations allow us to prove that

$$p^\star - p^{s+1} + \sum_{n=1}^N (d_{\theta_n}^{s+1})'\left[\rho_1 r_{n,z}^{s+1} + \rho_2 P' r_g^{s+1}\right] + - \sum_{n=1}^N \left[(\hat{\theta}_n^{s+1} - z_n^{s+1})'\delta_{n,1}^{s+1} + (P\hat{\theta}_n^{s+1} - \hat{\theta}^{g,s})'\delta_{n,2}^{s+1}\right] \geq 0. \tag{A.3}$$

By further adding and subtracting

$$\sum_{n=1}^N \left[(\hat{\theta}_n^{s+1} - z_n^{s+1})'\delta_{n,1}^\star + (P\hat{\theta}_n^{s+1} - \hat{\theta}^{g,s})'\delta_{n,2}^\star\right], \tag{A.4}$$

to (A.3), and by relying on the updates of the Lagrange multipliers in (9d)–(9e) and on the primal feasibility condition, a similar reasoning to the one exploited in Wei and Ozdaglar (2012) allows us to further show that

$$\begin{aligned}
&p^\star - p^{s+1} + \sum_{n=1}^N \left[\frac{\rho_1}{2}\|d_{z_n}^s\|_2^2 + \frac{\rho_2}{2}\|d_{\theta^g}^s\|_2^2\right] + \\
&- \sum_{n=1}^N \left[(\delta_{n,1}^\star)'(\hat{\theta}_n^{s+1} - z_n^{s+1}) + (\delta_{n,2}^\star)'(P\hat{\theta}_n^{s+1} - \hat{\theta}^{g,s+1})\right] + \\
&+ \sum_{n=1}^N \sum_{i=1}^2 \left[\frac{1}{2\rho_i}\|d_{\delta_{n,i}}^s\|_2^2\right] \geq \sum_{n=1}^N \left[\frac{\rho_1}{2}\|d_{z_n}^{s+1}\|_2^2 + \frac{\rho_2}{2}\|d_{\theta^g}^{s+1}\|_2^2\right] + \\
&+ \sum_{n=1}^N \left[\sum_{i=1}^2 \left[\frac{1}{2\rho_i}\|d_{\delta_{n,i}}^{s+1}\|_2^2\right] + \frac{\rho_1}{2}\|\hat{\theta}_n^{s+1} - z_n^s\|_2^2 + \right. \\
&\left. + \frac{\rho_2}{2}\|P\theta_n^{s+1} - \hat{\theta}^{g,s}\|_2^2\right],
\end{aligned} \tag{A.5}$$

with $d_{z_n}^{s+1}$ defined as in (A.2) and

$$d_{\delta_{n,i}}^s = \delta_{n,i}^s - \delta_{n,i}^\star, \quad i = 1, 2, \tag{A.6}$$

$$d_{\theta^g}^s = \hat{\theta}^{g,s} - \theta^{g\star}. \tag{A.7}$$

Since (A.5) holds for all $s = 0, 1, \ldots, k-1$, by summing over s and performing telescopic cancellations we obtain that

$$\begin{aligned}
&kp^\star - \sum_{s=0}^{k-1} p^{s+1} + \\
&- \sum_{n=1}^N \left[(\delta_{n,1}^\star)'\left(\sum_{s=0}^{k-1} \hat{\theta}_n^{s+1} - z_n^{s+1}\right) + (\delta_{n,2}^\star)'\left(\sum_{s=0}^{k-1} P\hat{\theta}_n^{s+1} - \hat{\theta}^{g,s+1}\right)\right] + \\
&+ \sum_{n=1}^N \left[\sum_{i=1}^2 \left(\frac{1}{2\rho_i}\|\delta_{n,i}^0 - \delta_{n,i}^\star\|_2^2\right) + \frac{\rho_1}{2}\|z_n^\star - z_n^0\|_2^2 + \right. \\
&\left. + \frac{\rho_2}{2}\|\theta^{g\star} - \hat{\theta}^{g,0}\|_2^2\right] \geq \sum_{n=1}^N \left[\sum_{i=1}^2 \left(\frac{1}{2\rho_i}\|d_{\delta_{n,1}}^k\|_2^2\right) + \right. \\
&+ \frac{\rho_1}{2}\|d_{z_n}^k\|_2^2 + \frac{\rho_2}{2}\|d_{\theta^g}^k\|_2^2 + \sum_{s=0}^{k-1} \frac{\rho_1}{2}\|\theta_n^{s+1} - z_n^s\|_2^2 + \\
&\left. + \sum_{s=0}^{k-1} \frac{\rho_2}{2}\|P\theta_n^{s+1} - \hat{\theta}^{g,s}\|_2^2\right].
\end{aligned} \tag{A.8}$$

The right-hand-side of the inequality in (26) easily follows from the definitions of the ergodic averages, the positivity of the right-hand-side of (A.8), the convexity of the cost function and primal feasibility.

Appendix B. Proof of Proposition 2

Similarly to Boyd et al. (2011), residuals and objective convergence is proven by showing that the weighted sum V^k, given by

$$V^k = \rho_2\|\hat{\theta}^{g,k} - \theta^{g\star}\|_2^2 + \frac{1}{N}\sum_{n=1}^N V_n^k,$$

$$V_n^k = \frac{1}{\rho_1}\|\delta_{n,1}^k - \delta_{n,1}^\star\|_2^2 + \rho_1\|z_n^k - z_n^\star\|_2^2 + \frac{1}{\rho_2}\|\delta_{n,2}^k - \delta_{n,2}^\star\|_2^2,$$

is a Lyapunov function for ADMM-RLS. This will be proven by showing that the following key inequalities hold:

$$\begin{aligned}
V^{k+1} \leq V^k - \frac{1}{N}\sum_{n=1}^N &\left[\rho_1\|r_{n,1}^{k+1}\|_2^2 + \rho_2\|r_{n,2}^{k+1}\|_2^2 + \right. \\
&\left. + \rho_1\|r_{n,z}^{k+1}\|_2^2 + \rho_2\|r_g^{k+1}\|_2^2\right],
\end{aligned} \tag{B.1}$$

$$\begin{aligned}
p^{k+1} - p^\star \leq -\sum_{n=1}^N &\left[(\delta_{n,1}^{k+1})' r_{n,1}^{k+1} + (\delta_{n,2}^{k+1})' r_{n,2}^{k+1} + \right. \\
&\left. + \rho_1(r_{n,z}^{k+1})'\zeta_{n,1}^{k+1} + \rho_2(r_g^{k+1})'\zeta_{n,2}^{k+1}\right],
\end{aligned} \tag{B.2}$$

$$p^\star - p^{k+1} \leq \sum_{n=1}^N \left[(\delta_{n,1}^\star)' r_{n,1}^{k+1} + (\delta_{n,2}^\star)' r_{n,2}^{k+1}\right], \tag{B.3}$$

with $r_{n,z}^{k+1}$ and r_g^{k+1} defined as in (21a) and (21b), respectively, and

$$\zeta_{n,1}^{k+1} = r_{n,1}^{k+1} + z_n^{k+1} - z_n^\star, \quad \zeta_{n,2}^{k+1} = r_{n,2}^{k+1} + \hat{\theta}^{g,k+1} - \theta^{g\star}.$$

Notice that objective convergence easily follows from (B.2)–(B.3). Instead, the inequality in (B.1) implies that V^k decreases at each iteration. Consequently, it holds that $V^k \leq V^0$. Since V^0 finite under Assumption 2, the variables $\theta^{g,k}$, $\{z_n^k, \delta_{n,1}^k, \delta_{n,2}^k\}_{n=1}^N$ have thus to be bounded. Based on (B.1) it can further be shown that

$$\sum_{k=0}^\infty \left\{ \rho_2\|r_g^{k+1}\|_2^2 + \frac{1}{N}\sum_{n=1}^N \left[\rho_1(\|r_{n,1}^{k+1}\|_2^2 + \|r_{n,z}^{k+1}\|_2^2) + \rho_2\|r_{n,2}^{k+1}\|_2^2\right] \right\} \leq V^0,$$

which implies residuals converge and that

$$r_{n,z}^{k+1} = z_n^{k+1} - z_n^k \xrightarrow[k\to\infty]{} 0, \quad n = 1, \ldots, N,$$

$$r_g^{k+1} = \hat{\theta}^{g,k+1} - \hat{\theta}^{g,k} \xrightarrow[k\to\infty]{} 0.$$

Note that, residual convergence and the boundedness of z_n^k and $\theta^{g,k}$ further result in the boundedness of the local estimates θ_n^k. Since the relation in (B.2) can be proven through straightforward manipulations of the result of Lemma 2, in the following we focus on the proofs of (B.1) and (B.3).

B.1. Proof of inequality (B.3)

Under Assumption 1, it holds that

$$\mathcal{L}^{\mathrm{o}}(\{\theta_n^\star, z_n^\star\}_{n=1}^N, \theta^{g\star}, \{\delta_{n,1}^\star, \delta_{n,2}^\star\}_{n=1}^N) \leq \mathcal{L}^{\mathrm{o}}(\{\hat{\theta}_n^{k+1}, z_n^{k+1}\}_{n=1}^N, \hat{\theta}^{g,k+1}, \{\delta_{n,1}^\star, \delta_{n,2}^\star\}_{n=1}^N).$$

Because of the primal feasibility of the saddle point, it further holds that

$$\theta_n^\star = z_n^\star, \quad P\theta_n^\star = \theta^{g\star} \quad \forall n \in \{1,\ldots,N\}.$$

By exploiting these relations and replacing $\mathcal{L}^{\mathrm{o}}$ with its definition, the proof of inequality (B.3) easily follows.

B.2. Proof of inequality (B.1)

Inequality (B.1) is proven via algebraic manipulations of the relationships in (B.2) and (B.3). By combining these inequalities and exploiting the updates in (9d)–(9e), it can be shown that the following inequality holds

$$\sum_{n=1}^N \Big[\frac{1}{\rho_1}(d_{\delta_{n,1}}^k)'(\delta_{n,1}^{k+1}-\delta_{n,1}^k) + \frac{1}{2\rho_1}\|\delta_{n,1}^{k+1}-\delta_{n,1}^k\|_2^2 + \\ +\frac{1}{\rho_2}(d_{\delta_{n,2}}^{k+1})'(\delta_{n,2}^{k+1}-\delta_{n,2}^k) + \frac{1}{2\rho_2}\|\delta_{n,2}^{k+1}-\delta_{n,2}^k\|_2^2 + \\ +\frac{\rho_1}{2}\|r_{n,1}^{k+1}\|_2^2 + \frac{\rho_2}{2}\|r_{n,2}^{k+1}\|_2^2 + \rho_1(r_{n,z}^{k+1})'(r_{n,1}^{k+1}+d_{z_n}^{k+1}) + \\ +\rho_2(r_g^{k+1})'(r_{n,2}^{k+1}+d_{\theta^g}^{k+1})\Big] \leq 0, \tag{B.4}$$

with $d_{z_n}^{k+1}$, $\{d_{\delta_{n,i}}^{k+1}\}_{i=1}^2$ and $d_{\theta^g}^{k+1}$ defined as in (A.2) and (A.6)–(A.7), respectively. Moreover, simple algebraic manipulations lead to the following equalities:

$$(d_{\delta_{n,i}}^{k+1})'(\delta_{n,i}^{k+1}-\delta_{n,i}^k) + \|\delta_{n,i}^{k+1}-\delta_{n,i}^k\|_2^2 = \frac{1}{2}\left[\|d_{\delta_{n,i}}^{k+1}\|_2^2 - \|d_{\delta_{n,i}}^k\|_2^2\right],$$

$$\frac{1}{2}\|r_{n,1}^{k+1}\|_2^2 + (r_{n,z}^{k+1})'(r_{n,1}^{k+1}+d_{z_n}^{k+1}) = \\ = \frac{1}{2}\left[\|r_{n,1}^{k+1}+r_{n,z}^{k+1}\|_2^2 + \|d_{z_n}^{k+1}\|_2^2 - \|d_{z_n}^k\|_2^2\right],$$

$$\frac{1}{2}\|r_{n,2}^{k+1}\|_2^2 + (r_g^{k+1})'(r_{n,2}^{k+1}+d_{\theta^g}^{k+1}) = \\ = \frac{1}{2}\left[\|r_{n,2}^{k+1}+r_g^{k+1}\|_2^2 + \|d_{\theta^g}^{k+1}\|_2^2 - \|d_{\theta^g}^k\|_2^2\right],$$

which hold for all $n \in \{1,\ldots,N\}$. Consequently, by using the definition of V^k, (B.4) can be equivalently recast as:

$$V^{k+1} - V^k \leq -\sum_{n=1}^N \left[\frac{\rho_1}{N}\|r_{n,1}^{k+1}+r_{n,z}^{k+1}\|_2^2 + \frac{\rho_2}{N}\|r_{n,2}^{k+1}+r_g^{k+1}\|_2^2\right].$$

To prove inequality (B.1), it is thus left to be shown that

$$\|r_{n,1}^{k+1}+r_{n,z}^{k+1}\|_2^2 \geq \|r_{n,1}^{k+1}\|_2^2 + \|r_{n,z}^{k+1}\|_2^2,$$

holds for $n = 1,\ldots,N$, and that

$$\sum_{n=1}^N \|r_{n,2}^{k+1}+r_g^{k+1}\|_2^2 \geq \sum_{n=1}^N \left[\|r_{n,2}^{k+1}\|_2^2 + \|r_g^{k+1}\|_2^2\right],$$

that can easily be verified through simple manipulations, by adopting the same reasoning as in Boyd et al. (2011), thus concluding the proof. □

Appendix C. Proof of Proposition 3

Let $V_n(k,t)$ be the Lyapunov-like function

$$V_n(k,t) = (\bar{\theta}_n^k(t)-\theta_n^{\mathrm{o}})'Q_n(t)(\bar{\theta}_n^k(t)-\theta_n^{\mathrm{o}}), \tag{C.1}$$

with $Q_n(t) = (\phi_n(t))^{-1}$. According to the definition of ergodic average $\bar{\theta}_n^k(t)$ in (25) and the decomposition of $\hat{\theta}_n^{s+1}(t)$ into the sum of $\hat{\theta}_n^{rls}(t)$ and $\hat{\theta}_n^{admm,s+1}(t)$, for $s = 0,1,\ldots,k-1$, it can be easily shown that the following holds:

$$\bar{\theta}_n^k(t) = \hat{\theta}_n^{rls}(t) + \frac{1}{k}\sum_{s=0}^{k-1}\hat{\theta}_n^{admm,s+1}(t).$$

By replacing $\bar{\theta}_n^k(t)$ in (C.1) with the previous definition and accounting for the equality in (13) and $\tilde{\theta}_n^{rls}(t) = \hat{\theta}_n^{rls}(t) - \theta_n^{\mathrm{o}}$, by properly completing the squares it can be shown that

$$V_n(k,t) = V_n^{rls}(t) + \left\|\tilde{\theta}_n^{rls} + \frac{1}{k}\sum_{s=0}^{k-1}\xi_n^s(t)\right\|_2^2 \\ + \frac{1}{k^2}\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right)'\phi_n(t)\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right) - (\tilde{\theta}_n^{rls}(t))'(\tilde{\theta}_n^{rls}(t)) \\ - \frac{1}{k^2}\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right)'\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right).$$

Since the two last terms are quadratic forms and are non-negative, it follows that

$$V_n(k,t) \leq V_n^{rls}(t) + \left\|\tilde{\theta}_n^{rls} + \frac{1}{k}\sum_{s=0}^{k-1}\xi_n^s(t)\right\|_2^2 + \\ + \frac{1}{k^2}\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right)'\phi_n(t)\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right).$$

According to Lemma 3 and from the properties of quadratic forms, it can be shown that

$$V_n(k,t) \leq \alpha\left\|\tilde{\theta}_n^{rls}(0)\right\|_2^2 + \sum_{\tau=1}^t (e_n(\tau))'e_n(\tau) + \\ + \left\|\tilde{\theta}_n^{rls} + \frac{1}{k}\sum_{s=0}^{k-1}\xi_n^s(t)\right\|_2^2 + \frac{1}{k^2}\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right)'\phi_n(t)\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right),$$

where α is the highest eigenvalue of $Q_n(0) = (\phi_n(0))^{-1}$. Consider now the fourth term in the inequality. By exploiting the properties of quadratic forms it holds that

$$\frac{1}{k^2}\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right)'\phi_n(t)\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right) \leq \beta(t)\left\|\frac{1}{k}\sum_{s=0}^{k-1}\xi_n^s(t)\right\|_2^2,$$

with $\beta(t)$ being the largest eigenvalue of $\phi_n(t)$. By adding and subtracting $\rho_1\theta_n^{\mathrm{o}} + \rho_2 P'P\theta_n^{\mathrm{o}}$ and

$$\frac{\rho_1}{k}\left(z_n^k(t) - \frac{\delta_{n,1}^k(t)}{\rho_1}\right) + \frac{\rho_2}{k}P'\left(\hat{\theta}^{g,k}(t) - \frac{\delta_{n,2}^k(t)}{\rho_2}\right)$$

within the squared 2-norm and, it follows that

$$\frac{1}{k^2}\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right)'\phi_n(t)\left(\sum_{s=0}^{k-1}\xi_n^s(t)\right) \leq \beta(t)(\varphi_n^k(\theta_n^{\mathrm{o}},t))^2, \tag{C.2}$$

with $\varphi_n^k(\theta_n^o, t)$ defined as in (31b). Moreover, by adding and subtracting

$$\frac{\rho_1}{k}\left(z_n^k(t) - \frac{\delta_{n,1}^k(t)}{\rho_1}\right) + \frac{\rho_2}{k}P'\left(\hat{\theta}^{g,k}(t) - \frac{\delta_{n,2}^k(t)}{\rho_2}\right) + \\ + \rho_1(\hat{\theta}_n^{rls}(t) - \theta_n^o) + \rho_2 P'(P\hat{\theta}_n^{rls}(t) - \theta^{g,o}),$$

within the norm in the third term, straightforward manipulation leads to

$$\tilde{\theta}_n^{rls}(t) + \frac{1}{k}\sum_{s=0}^{k-1}\xi_n^s(t) = \mathcal{Q}\tilde{\theta}_n^{rls}(t) + \varphi_n^k(\hat{\theta}_n^{rls}(t), t), \tag{C.3}$$

with $\mathcal{Q}$ and $\varphi_n^k(\hat{\theta}_n^{rls}(t), t)$ defined as in Eqs. (31a) and (31c), respectively. The RLS updates in (15)–(17) can then be iteratively exploited to decompose $\tilde{\theta}_n^{rls}(t)$ as

$$\tilde{\theta}_n^{rls}(t) = \phi_n(t)\left[\mathcal{Q}_n(0)\tilde{\theta}_n^{rls}(0) + \sum_{\tau=1}^{t-1}\mathcal{Q}_n(\tau)e_n(\tau)\right] + \\ + K_n(t)e_n(t). \tag{C.4}$$

By substituting (C.4) into (C.3), and then exploiting the triangular inequality, it can be shown that the following holds:

$$V_n(k,t) \leq \alpha\left\|\tilde{\theta}_n^{rls}(0)\right\|_2^2 + \sum_{\tau=1}^{t}(e_n(\tau))'e_n(\tau) + \\ + \left[\eta_n(t) + \varphi_n^k(\hat{\theta}_n^{rls}(t), t)\right]^2 + \beta(t)(\varphi_n^k(\theta_n^o, t))^2.$$

For the properties of quadratic forms, it further holds that $V_n(k,t) \geq \nu(t)\left\|\bar{\theta}_n^k(t) - \theta_n^o\right\|_2^2$, where $\nu(t) > 0$ is the highest eigenvalue of $\mathcal{Q}_n(t)$. By exploiting this last relationship, the inequality in (30) easily follows. □

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.

B., Lindoff, & J., Holst (1999). Convergence analysis of the RLS identification algorithm with exponential forgetting in stationary ARX-structures. *International Journal of Adaptive Control and Signal Processing*, *13*(1), 1–22.

Benning, M., Knoll, F., Schönlieb, C., & Valkonen, T. (2015). Preconditioned ADMM with nonlinear operator constraint. https://arxiv.org/abs/1511.00425.

Bittanti, S., Bolzern, P., & Campi, M. (1990). Recursive least-squares identification algorithms with incomplete excitation: convergence analysis and application to adaptive control. *IEEE Transactions on Automatic Control*, *35*(12), 1371–1373.

Boem, F., Ferrari, R. M. G., Keliris, C., Parisini, T., & Polycarpou, M. M. (2017). A distributed networked approach for fault detection of large-scale systems. *IEEE Transactions on Automatic Control*, *62*(1), 18–33.

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, *3*(1), 1–122.

Breschi, V., Kolmanovsky, I., & Bemporad, A. (2018). Cloud-aided collaborative estimation by ADMM-RLS algorithms for connected vehicle prognostics. In *2018 Annual American control conference* (pp. 2727–2732).

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). Diffusion LMS strategies for distributed estimation. *IEEE Transactions on Signal Processing*, *58*(3), 1035–1048.

Forero, P. A., Cano, A., & Giannakis, G. B. (2010). Consensus-based distributed support vector machines. *Journal of Machine Learning Research (JMLR)*, *11*, 1663–1707.

Howell, M. N., Whaite, J. P., Amatyakul, P., Chin, Y. K., Salman, M. A., Yen, C. H., et al. (2010). Brake pad prognosis system. US Patent 7, 694, 555. https://www.google.com/patents/US7694555.

Kar, S., & Moura, J. M. F. (2013). Consensus + innovations distributed inference over networks: cooperation and sensing in networked systems. *IEEE Signal Processing Magazine*, *30*(3), 99–109.

Kidambi, N., Harne, R. L., Fujii, Y., Pietron, G. M., & Wang, K. W. (2014). Methods in vehicle mass and road grade estimation. *SAE International Journal of Passenger Cars-Mechanical Systems*, *7*(2014-01-0111), 981–991.

Li, Z., Kolmanovsky, I., Atkins, E. M., Lu, J., Filev, D. P., & Bai, Y. (2017). Road disturbance estimation and cloud-aided comfort-based route planning. *IEEE Transactions on Cybernetics*, *PP*(99), 1–13. http://dx.doi.org/10.1109/TCYB.2016.2587673.

Li, Z., Kolmanovsky, I., Atkins, E., Lu, J., Filev, D. P., & Michelini, J. (2016). Road risk modeling and cloud-aided safety-based route planning. *IEEE Transactions on Cybernetics*, *46*(11), 2473–2483. http://dx.doi.org/10.1109/TCYB.2015.2478698.

Ljung, L. (1999). *System identification: theory for the user*. NJ: Prentice-Hall Englewood Cliffs.

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

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.

Mell, P. M., & Grance, T. (2011). *SP 800-145. The NIST definition of cloud computing*: *Technical Report*, Gaithersburg, MD, United States: National Institute of Standards & Technology.

Nassif, R., Vlaski, S., & Sayed, A. H. (2019). Adaptation and learning over networks under subspace constraints – Part I: Stability analysis.

Olfati-Saber, R. (2007). Distributed Kalman filtering for sensor networks. In *2007 46th IEEE conference on decision and control* (pp. 5492–5498). Dec 2007.

Ozatay, E., Onori, S., Wollaeger, J., Ozguner, U., Rizzoni, G., Filev, D. P., et al. (2014). Cloud-based velocity profile optimization for everyday driving: A dynamic-programming-based solution. *IEEE Transactions on Intelligent Transportation Systems*, *15*(6), 2491–2505.

Rajamani, R. (2012). *Vehicle dynamics and control* (2nd ed.). Springer Science.

Ram, S. S., Nedić, A., & Veeravalli, V. V. (2007). Stochastic incremental gradient descent for estimation in sensor networks. In *2007 conference record of the forty-first asilomar conference on signals, systems and computers* (pp. 582–586). Nov 2007.

Sahu, A. K., Jakovetić, D., & Kar, S. (2018). $\mathcal{CIRFE}$: A distributed random fields estimator. *IEEE Transactions on Signal Processing*, *66*(18), 4980–4995.

Sahu, A. K., Kar, S., Moura, J. M. F., & Poor, H. V. (2016). Distributed constrained recursive nonlinear least-squares estimation: Algorithms and asymptotics. *IEEE Transactions on Signal and Information Processing over Networks*, *2*(4), 426–441.

Sayed, A. H., & Lopes, C. G. (2006). Distributed recursive least-squares strategies over adaptive networks. In *2006 fortieth asilomar conference on signals, systems and computers* (pp. 233–237). Oct 2006.

Schizas, I. D., Mateos, G., & Giannakis, G. B. (2007). Consensus-based distributed recursive least-squares estimation using ad hoc wireless sensor networks. In *2007 conference record of the forty-first asilomar conference on signals, systems and computers* (pp. 386–390). Nov 2007.

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.

Taheri, E., Gusikhin, O., & Kolmanovsky, I. (2016). Failure prognostics for in-tank fuel pumps of the returnless fuel systems. In *Dynamic systems and control conference*. Oct 2016.

Vahidi, A., & Eskandarian, A. (2003). Research advances in intelligent collision avoidance and adaptive cruise control. *IEEE Transactions on Intelligent Transportation Systems*, *4*(3), 143–153.

Vahidi, A., Stefanopoulou, A., & Peng, H. (2005). Recursive least squares with forgetting for online estimation of vehicle mass and road grade: theory and experiments. *Vehicle System Dynamics*, *43*(1), 31–55.

Valkonen, T. (2014). A primal–dual hybrid gradient method for nonlinear operators with applications to MRI. *Inverse Problems*, *30*(5), 1–45.

Wei, E., & Ozdaglar, A. (2012). Distributed alternating direction method of multipliers. In *2012 IEEE 51st IEEE conference on decision and control* (pp. 5445–5450). Dec 2012.

Valentina Breschi received her Master's degree in Electrical and System Engineering in 2014 from the University of Florence, Italy. She received her Ph.D. in Control Engineering in 2018 from IMT School for Advanced Studies Lucca. In 2017 she was a visting scholar at the Department of Aerospace Engineering, University of Michigan, Ann Arbor. In 2018–2020 she held a postdoctoral position at Politecnico di Milano, where she is currently a junior Assistant Professor. Her main research interests include systems identification, data-driven control, data-analysis and policy design for mobility systems.

Alberto Bemporad received his Master's degree in Electrical Engineering in 1993 and his Ph.D. in Control Engineering in 1997 from the University of Florence, Italy. In 1996/97 he was with the Center for Robotics and Automation, Department of Systems Science & Mathematics, Washington University, St. Louis. In 1997–1999 he held a postdoctoral position at the Automatic Control Laboratory, ETH Zurich, Switzerland, where he collaborated as a senior researcher until 2002. In 1999–2009 he was with the Department of Information Engineering of the University of Siena, Italy, becoming an Associate Professor in 2005. In 2010–2011 he was with the Department of Mechanical and Structural Engineering of the University of Trento, Italy. Since 2011 he is Full Professor at the IMT School for Advanced Studies Lucca, Italy, where he served as the Director of the institute in 2012–2015. He spent visiting periods at Stanford University, University of Michigan, and Zhejiang University. In 2011 he cofounded ODYS S.r.l., a company specialized in developing model predictive control systems for industrial production. He has published more than 350 papers in the areas of model predictive control, hybrid systems, optimization, automotive control, and is the co-inventor of 16 patents. He is author or coauthor of various software packages for model predictive control design and implementation, including the Model Predictive Control Toolbox (The Mathworks, Inc.) and the Hybrid Toolbox for MATLAB. He was an Associate Editor of the IEEE Transactions on Automatic Control during 2001–2004 and Chair of the Technical Committee on Hybrid Systems of the IEEE Control Systems Society in 2002–2010. He received the IFAC High-Impact Paper Award for the 2011–14 triennial and the IEEE CSS Transition to Practice Award in 2019. He is an IEEE Fellow since 2010.

Ilya V. Kolmanovsky is a professor in the department of aerospace engineering at the University of Michigan. His research interests are in control theory for systems with state and control constraints, and in control applications to aerospace and automotive systems. He received his Ph.D. degree from the University of Michigan in 1995. He is a Fellow of IEEE and is named as an inventor on over a hundred of United States patents.