52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Distributed Extremum Seeking Control of Networked Large-scale Systems Under Constraints

Jin-Ming Xu[†] and Yeng Chai Soh[†]

Abstract— **We consider the distributed optimization problem of networked large-scale systems under specified constraints. In particular, we aim to optimize the steady-state performance of the system by making it (in a distributed manner) working at the optimal operation point. To solve the problem, a primal-dual method is employed to seek the optimal setpoint in conjunction with Extremum Seeking Control (ESC) which is utilized as a tool for efficient gradient estimation without knowing the specific form of the cost functions as well as the constraints of the system. The proposed overall scheme is termed as Distributed Extremum Seeking Control (D-ESC) as it is designed based on the ESC scheme and implemented in a distributed way. It will be shown, by resorting to singular perturbation and averaging theory as well as duality theory, that the overall networked large-scale system equipped with D-ESC is Semi-globally Practically Asymptotically Stable (SPAS) and thus will eventually converge to the neighborhood of the Pareto-optimal solution of the primal problem.**

I. Introduction

Most systems encountered in reality, as they become more integrated, can and should be captured as networked large-scale interconnected systems consisting of a great number of subsystems coupling through either physical interactions or specified global constraints, e.g., plant-wide processes [22], power grids [1], wind farms [15], and multi-robot systems [20], just to name a few. The control of such systems is extremely difficult due to its large-scale property and global constraints which probably render the conventional ways inefficient or even impossible.

To deal with such systems, distributed control is renewed nowadays with emphasis on the local communication among subsystems and local control of individual subsystems [4]. At the heart of this kind of control is the distributed optimal control which not only stabilizes the overall system in a distributed fashion but also, most importantly, optimizes its *transient* as well as *steady-state* performance. A well established one is that of distributed model predictive control(D-MPC) [22, 8, 11, 7]. This control method mainly focuses on real-time optimization of the *transient* performance of the system with known setpoint in a distributed way. In contrast to this method, there is also a resurgence of interest in extremum seeking control(ESC) which, instead, attempts to optimize the *steady-state* performance of the system in real-time without knowing the specific function of performance so long as its value can be measured [2, 28].

In addition, it is well known that game theory, as a modeling technique dealing with the optimization problem of multiple decision makers, is closely correlated with distributed optimal control [5]. In particular, Waslander et al. have made an attempt to solve the decentralized optimization problem by utilizing Nash Bargaining method [33]. Semsar-Kazerooni and Khorasani have considered the multiple LQR problem from the viewpoint of game theory and attempted to obtain the Pareto-efficient solution [25]. On the other hand, ESC has been employed and implemented in a distributed manner for seeking Nash equilibrium [27, 10], but it is still limited to problems without constraints and the solution may not be Pareto-efficient [5].

Moreover, dual decomposition, as an effective way to relaxing constraints or couplings, appears to be a good candidate for distributed optimization of large-scale systems mostly due to its Pareto-efficient solution as well as its ability to be implemented in a distributed way. With this method, existing applications include distributed estimation [24], Multi-agent Optimization [31], distributed optimal control [32, 21], resource allocation in wireless communication [34] and network utility maximization [14] and so on. However, the capacity of dual decomposition relies on the efficiency of evaluating the subgradient of the Lagrange dual function, e.g., the objective function should be separable. Recently, Nedic and Ozdaglar [16] studied the distributed optimization problem using consensus theory as proposed in [19], wherein each agent is trying to optimize its own objective while reaching consensus with the others. Zhu et al [36] and Yuan et al. [35] extended this method taking into account constraints by utilizing primal-dual subgradient method with projection. Again, they require the exact form of the cost functions and constraints associated with the system.

In this paper, we propose a new control scheme, termed as D-ESC, to solve the distributed constraint optimization problem associated with the steady-state performance of networked large-scale systems. In particular, we use the primal-dual method [9, 17] to solve the problem in a distributed way with the gradient estimated by ESC without knowing the specific form of the cost and constraint functions(*Note that we do not require the function to be separable*). The proposed scheme is very important in practice in that it is inherently adaptive and robust as no specific form of the function of performance is required. For instance, in wind farms, the analytical form of the power generated is not accessible to the turbines due to their complex aerodynamic interaction. However, with ESC scheme being involved, central to this proposed method is that of the property of convergence, especially when the system is subjecting to global constraints. On the convergence issues of ESC, Krstic

[†]The authors are with School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. xuji0016@e.ntu.edu.sg; eycsoh@ntu.edu.sg

and Wang [13] first gave a rigorous proof on the local stability of the general form of ESC using averaging and singular perturbation theory. For extension, Tan et al. [29] considered and proved the non-local stability of ESC by making stronger assumption on the static-map function and assuming that the controller parameters are properly tuned. However, in the context of ESC, there is little literature devoted to the optimization problem under constraints. In this work, by using singular perturbation and averaging theory [12] and duality theory [6], we will prove that the overall networked system equipped with the D-ESC scheme is semi-globally practically asymptotically stable (SPAS) [29] and thus will eventually converge to the neighborhood of the Pareto-optimal solution of the primal problem.

II. Preliminaries

A continuous function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{K}$ if it is strictly increasing and $\alpha(0) = 0$. It is of class $\mathcal{K}_\infty$ if unbounded and $\alpha(r) \to \infty$ as $r \to \infty$. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to belong to class $\mathcal{KL}$ if $\beta(\cdot, t)$ is of class $\mathcal{K}$ for each $t \geq 0$ and $\beta(r, \cdot)$ is decreasing to zero for each $r \geq 0$. In addition, we denote by x_i the ith component of a vector x. We will use $\{x_i\}$ to represent the proper collection of all $*_i,\ i \in \mathcal{I}$ and $O(\cdot)$ to denote the order of magnitude. Whenever it is clear, we will suppress some arguments of certain function for brevity. Throughout the paper, the concept of convergence and stability is used for algorithms and dynamic systems respectively.

Definition 1 (SPAS [18, 29]): Consider a parameterized system $\dot{x} = f(x, \varepsilon)$, which is continuous and locally Lypschitz in (x, ε) and let (Δ, δ) be a pair of strictly positive numbers. Then, the system is said to be semi-globally practically asymptotically stable (SPAS) if there exists real numbers $\varepsilon^* = \varepsilon^*(\Delta, \delta) > 0$ such that for any $x_0 \in \Delta$ and for each $\varepsilon \in (0, \varepsilon^*)$, we have $|x(t)| \leq \beta(|x_0|, t) + \delta$.

III. Problem Formulation

A. Modeling of Networked Large-scale Systems

In this section, the model of networked large-scale interconnected systems will be given based on the analysis of its topology of interaction. In particular, similar to that of [8], we consider a large-scale system consisting of N subsystems. For any subsystem $j,\ j \in \mathcal{I} := \{1, 2, ..., N\}$ having any component of its dynamic state coupled into subsystem $i \neq j$ through cost/output functions, it will be regarded as the upstream neighbor of subsystem i, denoted as $\mathcal{N}_i^u$. Correspondingly, subsystem i will be referred to as the downstream neighbor of subsystem j, denoted as $\mathcal{N}_i^d$. Having this in mind, we present the model of the dynamic system with its associated global constraints[1] as follows:

$$\begin{cases} \dot{x}_i = f_i(x_i, x_{-i}^l, u_i) \\ y_i = h_i(x_i, x_{-i}^l, u_i),\ i \in \mathcal{I} \\ \quad s.t.\ g(x, u) \leq 0 \end{cases} \tag{1}$$

where $x_i \in \mathcal{D}_i \subset \mathcal{R}^{n_i}$ is the dynamic state of subsystem i, while $x_{-i}^l = \{x_j\}, j \in \mathcal{N}_i^u$ denotes the dynamic state of the upstream neighbors of subsystem i, $u_i \in \Gamma_i \subset \mathcal{R}^{m_i}$ is the control input, $y_i \in \mathcal{R}$ a scalar representing the cost induced by the operation of the system, e.g., dissipative energy. $x \in \mathcal{R}^n$ and $u \in \mathcal{R}^m$ is the state and input of the whole system. While the cost/output function h_i and the global constraints g could be unknown so long as they can be measured, it should be noted that the dynamic function f_i is known a priori and assumed to be continuous and locally Lipschitz in (x_i, u_i), uniformly in x_{-i}. In addition, we assume that the cost/output function h_i is only known to the subsystem i while the global constraints g can be somehow measured by all the subsystems involved.

Moreover, it is also assumed that each subsystem can only exchange information with its neighbors due to limitation of communication. While some issues are involved in real communication, e.g., packet loss, quantization and delay, we assume perfect communication among subsystems.

B. A Primal Optimization Problem

Given the above dynamic system (1), our objective is to design a distributed feedback control scheme so as to make the system working at the optimal operation point while subjecting to some specified constraints. As a good candidate, extremum seeking control (ESC) [2] will be employed for fulfilling this task. In [13, 29], it has been shown that using ESC scheme will give rise to a singularly perturbed system with the two-time-scale property, the boundary-layer system[2] associated with the fast time-scale and the reduced(quasi-steady-state) system associated with the relatively slower time-scale [12]. A standard procedure for the analysis of the stability of the system is the two-time-scale decomposition. That is, we can analyze the stability of the boundary-layer system and the reduced system separately and then investigate the stability of the overall system based on singular perturbation theory. Thus, by freezing the boundary-layer system at its equilibrium to obtain the reduced system, we can rewrite the output function y_i and the global constraints g system (1) in a quasi-steady-state form as follows[3]:

$$\bar{y}_i = \bar{h}_i(\bar{x}_i, \bar{x}_{-i}),\ i \in \mathcal{I} \quad s.t.\ \bar{g}(\bar{x}) \leq 0 \tag{2}$$

where $\bar{x}_i \in \mathcal{C}_i \subset \mathcal{R}^{m_i}$ with $\mathcal{C}_i$ being the local convex constraint set, and $\bar{x}_{-i}^l = \{\bar{x}_j\}, j \in \mathcal{N}_i^u$ denote the properly selected states of subsystem i and its neighbors respectively (the reader is referred to [26] for the detail on how to select controlled variables), which are to be controlled and need to satisfy $f_i(x_i^{ss}, x_{-i}^{ss}, u_i^{ss}) = 0$, where $x_i^{ss} = [\bar{x}_i, \bar{x}_i^r]^T$ is the steady-state with $\bar{x}_i^r$ as the one that do not directly contribute the output and the constraint. It worth mentioning that $\bar{x}_i$ should have the same dimension as the input of the subsystem i, which represents the degree-of-freedom available for optimizing the system operation.

[1] For the case of equality constraints $g_i(x_i, x_{-i}, u_i) = 0$, we could rewrite it as $g_i(x_i, x_{-i}, u_i) \leq 0$ and $-g_i(x_i, x_{-i}, u_i) \leq 0$.

[2] The boundary-layer system here refers to the physical dynamic system that is required to track certain specified setpoint.

[3] In the sequel, $\bar{*}$ stands for the state or function related to the reduced model with $*$ as the real one.

To guarantee the stability of the overall system, we first make the following assumption:

Assumption 1: There exists a distributed control law $u^* \in \Gamma$ which is preferred to be optimal, e.g., D-MPC, such that for any given feasible[4] setpoint $\bar{x}^s \in \{\mathcal{C}_i\} \subset \mathcal{R}^m$, the system (1) can be asymptotically driven to it in such a way that is fast enough as compared with that of D-ESC(which will be presented in the following section).

Then, we consider the following minimization problem associated with the reduced system subjecting to the global quasi-steady-state constraints:

$$\begin{aligned} \bar{x}^s = \underset{\bar{x}\in\mathcal{R}^m}{\operatorname{argmin}}\ J := \sum_{i=1}^{N} \bar{h}_i(\bar{x}_i, \bar{x}_{-i}), \\ s.t.\ \bar{x}_i \in \mathcal{C}_i,\ i \in \mathcal{I}, \text{and } \bar{g}(\bar{x}) \leq 0 \end{aligned} \tag{3}$$

where $J_i = \bar{h}_i(\bar{x}_i, \bar{x}_{-i})$ denotes the induced individual cost of subsystem i in the quasi-steady-state form and x^s the particular state or setpoint attaining the minimum of the overall cost function J, while $\bar{g}$ is the global quasi-steady-state constraints with which the system should end up.

C. Dual Optimization Problem with Decomposition

By using the duality theory [6], the corresponding dual optimization problem of (3) can be described as follows:

$$\begin{cases} \Phi := \sum_{\forall i\in\mathcal{I}} \bar{h}_i(\bar{x}_i, \bar{x}^l_{-i}) + \lambda^T \bar{g}(\bar{x}) \\ D^* := \max_{\lambda\geq 0} Q(\lambda), \quad Q := \inf_{\bar{x}\in\mathcal{C}} \Phi(\bar{x}, \lambda) \end{cases} \tag{4}$$

where Φ is the Lagrangian function, D^* is the overall dual optimal value, Q is the corresponding Lagrange dual function, $\bar{x}$ denotes the collection of the selected states of the overall system, λ the introduced dual variables accounting for the global constraints $\bar{g}$ imposing on the system.

In order to solve the optimization problem in a distributed manner, we define the following sub-dual problem for each subsystem i:

$$\begin{cases} \bar{\phi}_i := \bar{h}_i(\bar{x}_i, \bar{x}^l_{-i}) + \bar{\Delta}_i \\ \bar{\Delta}_i := \sum_{\forall j\in\mathcal{N}^d_i} \bar{h}_j(\bar{x}_j, \bar{x}^l_{-j}) + \lambda_i^T \bar{g}(\bar{x}). \\ d_i^* := \max_{\lambda_i\geq 0} q_i(\lambda_i), \quad q_i := \inf_{\bar{x}_i\in\mathcal{C}_i} \bar{\phi}_i(\bar{x}_i, \bar{x}^g_{-i}, \lambda_i) \end{cases} \tag{5}$$

such that for $\forall \bar{x}_i \in \mathcal{C}_i$, we have

$$\frac{\partial \bar{\phi}_i(\bar{x}_i, \bar{x}^g_{-i}, \lambda_i)}{\partial \bar{x}_i} = \frac{\partial \Phi(\bar{x}, \lambda_i)}{\partial \bar{x}_i} \tag{6}$$

where $\bar{\phi}_i(\bar{x}_i, \bar{x}^g_{-i}, \lambda)$ is the Lagrangian function of the subsystem i, d_i^* is the dual optimal value, q_i the corresponding Lagrange dual function, $\bar{x}^g_{-i}$ denotes the collection of the selected states of all the subsystems but i, λ_i the dual variables introduced by subsystem i to account for the global constraints and Δ_i the information required by subsystem i for implementing the individual ESC algorithm.

[4]To be feasible means the setpoint should satisfy the reduced (quasi-steady-state) model.

In order to guarantee zero duality gap between the primal problem and the dual one, we make the following assumption on the Lagrangian function Φ.

Assumption 2: The primal optimization problem (3) has an optimal solution $J^* > -\infty$. All cost/output functions and constraints are differentiable and convex for $\forall \bar{x} \in \mathcal{C}$ with the domain $\mathcal{C}$ being a convex and campact set [6].

Remark 1: Assumption 2 could be further relaxed so long as the ESC algorithm as will be introduced in the following section is capable of seeking the global extremum in the presence of local extremum [30].

IV. A PRIMAL-DUAL APPROACH BASED ON ESC

A. The Primal-Dual Gradient System

The Primal-Dual method, originated from the seminal work of [3], has been recently introduced to solve the distributed optimization problem (such as the optimization problem (3)) [9, 17, 21, 35]. It can be described in the generalized form as follows:

$$\begin{cases} \dot{x} = P_{T_x\mathcal{D}}[-\Gamma_x \cdot \nabla_x \Phi(x, \lambda)] \\ \dot{\lambda} = P_{T_x\mathcal{R}^+}[\Gamma_\lambda \cdot \nabla_\lambda \Phi(x, \lambda)] \end{cases} \tag{7}$$

where Γ_x and Γ_λ are positive diagonal matrices, which could be regarded as the constant stepsize in discrete case, ∇ is the differential operator, $\Phi(x, \lambda)$ denotes the Lagrangian function and $P_{T_x\mathcal{M}}[w]$ denotes the projection of w on the tangent space of $\mathcal{M}$, or the projection of x on $\mathcal{M}$ with the dynamic governed by $\dot{x} = w$, and thus by $P_{\mathcal{D}}[x]$ and $P_{\mathcal{R}^+}[x]$ we mean the projection of x and λ on $\mathcal{D}$ and $\mathcal{R}^+$ respectively, where $\mathcal{R}^+ := \{x | x \geq 0\}$. Note that in real practice, the projection is used to deal with hard constraints which can not be violated at any time, e.g., min-max constraints.

For convex hard constraints, the projection is defined as follows:

$$P_{T_x\mathcal{M}}[w] = \begin{cases} \lim_{\tau\to 0^+} \frac{P_{\mathcal{M}}(x+\tau w) - P_{\mathcal{M}}(x)}{\tau},\ x \in \partial\mathcal{M} \\ w,\ x \in \mathcal{M}, \end{cases} \tag{8}$$

where $\mathcal{M}$ is a closed convex set while $\partial\mathcal{M}$ and $T_x\mathcal{M}$ denote the boundary and the tangent space of $\mathcal{M}$ respectively.

With the above definition, we have the following property of the projection operator:

$$(x - x^*)^T P_{T_x\mathcal{M}}[w] \leq (x - x^*)^T w \quad \forall x,\ x^* \in \mathcal{M} \subset \mathcal{R}^m. \tag{9}$$

Proof: Let x and $x^* \in \mathcal{M} \subset \mathcal{R}^m$, where $\mathcal{M}$ is a closed convex set. Then, with the definition of the projection operator and let $\eta = x - x^*$, we have

$$\begin{aligned} (x - x^*)^T P_{T_x\mathcal{M}}[w] &= \lim_{\tau\to 0^+} \frac{\eta^T [P_{\mathcal{M}}(x + \tau w) - P_{\mathcal{M}}(x)]}{\tau} \\ &\leq \lim_{\tau\to 0^+} \frac{(\eta + \tau w)^T [(x + \tau w) - x]}{\tau} \\ &- \lim_{\tau\to 0^+} \frac{\tau w^T [P_{\mathcal{M}}(x + \tau w) - P_{\mathcal{M}}(x)]}{\tau} \\ &= \eta^T w - 0 = (x - x^*)^T w \end{aligned}$$

where we have used the non-expansive property of projection [6]. That is, for $\forall x \in \mathcal{M}$, we have

$$[(x^* - P_{\mathcal{M}}(x + \tau w)]^T [(x + \tau w) - P_{\mathcal{M}}(x + \tau w)] \le 0.$$

With $-\|[(x + \tau w) - P_{\mathcal{M}}(x + \tau w)]\|_2^2 \le 0$, we obtain

$$[(x^* - (x + \tau w)]^T [(x + \tau w) - P_{\mathcal{M}}(x + \tau w)] \le 0,$$

leading to

$$(\eta + \tau w)^T P_{\mathcal{M}}(x + \tau w) \le (\eta + \tau w)^T (x + \tau w),$$

which completes the proof. ■

B. ESC for Each Subsystem

In order to obtain the gradient information[5] of $\nabla_x \phi(x, \lambda)$ for the gradient system (7), ESC scheme is employed to do the estimation. The reader is referred to [2] for details concerning the general ESC scheme. As shown in Fig. 1, the extremum seeking control scheme has as its input the output of the subsystem i and the external coupling information Δ_i as defined in (5) and, based on that, manage to estimate the optimal setpoint as the reference signal for the system to track. It should be noted however that, as indicated or implied in [13, 28], the high-pass and low-pass filters involved in the scheme can be actually removed without impacting on the stability of the system since the objective of using them is to increase the system dynamic performance which is not the focus of this paper, e.g., attenuating the output oscillation. As such, in the stability analysis, we only consider the first-order ESC and it is not difficult as shown in [29] to use the same techniques to obtain the same results for higher order ESC which take into account the high-pass and low-pass filters. From Fig. 1, again, by fixing the boundary-layer system at its equilibrium, we have the dynamic system of ESC in the quasi-steady-state form as follows:

$$\dot{\hat{x}}_i = P_{\mathcal{C}_i}[-\varepsilon \phi_i(\hat{x}_i + a \cdot \sin(w_i t), \bar{x}^g_{-i}, \lambda_i) \cdot \sin(w_i t)]. \quad (10)$$

where $\hat{x}$ is the estimation of the optimal setpoint and $w_i = [w_{i1}, \ldots, w_{ij}, \cdots, w_{im_i}]^T$ is the column vector of the frequency of the dither signal added to the subsystem i and ε, a is some small positive number.

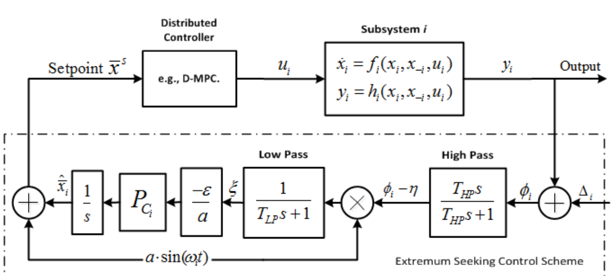

Fig. 1. The ESC scheme for subsystem i.

C. The D-ESC Scheme

Combing the above two dynamic systems (7) and (10), we obtain the overall dynamic system in the quasi-steady-state form, termed as Distributed Extremum Seeking Control (D-ESC) scheme, as follows:

$$\begin{cases} \dot{\hat{x}}_i = P_{\mathcal{C}_i}[-\varepsilon \phi_i(\hat{x}_i + a \cdot \sin(w_i t), \bar{x}^g_{-i}, \lambda_i) \\ \qquad \cdot \sin(w_i t)], i \in \mathcal{I} := \{1, 2, ..., N\} \\ \dot{\lambda}_i = P_{\mathcal{R}^+}[\Gamma_{\lambda_i} \cdot \nabla_{\lambda_i} \Phi(\hat{x}, \lambda_i)] \end{cases} \quad (11)$$

D. Overall Algorithm for D-ESC

In this section, we will summarize the above overall control scheme by proposing for each subsystem i the corresponding algorithm as given in *Algorithm* 1. In addition, the dual variables are assumed to be initialized with the same value and updated synchronously from then on. Thus, we have $\lambda_i = \lambda$, $\forall i \in \mathcal{I}$ all the time. In the sequel, we will drop its subscript i for brevity. In fact, for asynchronous cases, we can resort to consensus theory as in [16, 36, 35], replacing the dual update law with

$$\dot{\lambda}_i = P_{\mathcal{R}^+}[\sum_{j \in \mathcal{N}_i} (\lambda_j - \lambda_i) + \Gamma_{\lambda_i} \cdot \nabla_{\lambda_i} \Phi(\hat{x}, \lambda_i)],$$

which is more practical in real situation.

Algorithm 1 Subsystem i

1) Initialize the dual variable λ and the estimation of the initial state of subsystem i.
2) Receive from the neighbors $j \in \mathcal{N}_i$ the data of h_j and obtain the value of h_i and g by measurement.
3) Update λ and the estimated optimal setpoint $\hat{x}_i$ according to the scheme (11).
4) Communicate the value of h_i to the neighbors.
5) Repeat step (2)-(4) at each specified time interval.

V. Convergence Analysis

In this section, we will investigate the convergence property of the proposed D-ESC algorithm as well as the stability of the overall networked dynamic system. As mentioned early on, the low-pass and high-pass filters in the loop of ESC is not essential for analyzing the convergence of the proposed algorithm. To this end, we will give the convergence analysis based on the simplified version, the first-order ESC scheme which eliminates the filters involved in the control loop.

A. Averaging Analysis of Multi-variable ESC Scheme

In [23], Rotea has presented a simple averaged model for stability and performance analysis of multi-variable ESC scheme in an attempt to provide guidelines for properly tuning its parameters. In line with this work, we attempt to utilize the averaging theory [12] to examine the capability of gradient estimation of ESC[6].

Lemma 1 (Approximate Gradient Estimation of ESC): Consider the dynamic system of ESC described by (10)

[5] We only consider the gradient estimation of $\nabla_x \phi(x, \lambda)$ since it is much easy to obtain the gradient $\nabla_\lambda \phi(x, \lambda)$.

[6] We only consider first order approximation though high-order approximation is available when certain additional orthogonality conditions hold.

without projection. Assume that *Assumption* 2 holds and $w_{ij} \neq w_{ik}, \forall i \in \mathcal{I}, j,k \in \mathcal{N}_{x_i} := \{1,\ldots,j,\ldots,m_i\}$. Then, there exist $\bar{\varepsilon}$ and $\bar{a}$ such that for $\forall \varepsilon \in (0,\bar{\varepsilon})$ and $\forall a \in (0,\bar{a})$, system (10) can be rewritten in the averaged form as:

$$\dot{\hat{\bar{x}}}_i^{av} = -\frac{\varepsilon a}{2}\nabla_{\hat{\bar{x}}_i^{av}}\bar{\phi}_i(\hat{\bar{x}}_i^{av}, \hat{\bar{x}}_{-i}^g, \lambda). \tag{12}$$

with the approximation error as $O(\varepsilon a^2 + \varepsilon^2)$.

Proof: Let $\bar{w} = gcd(\{w_{ij}\})$[7], $T = \frac{2\pi}{\bar{w}}$ be the frequency and period of the overall system respectively. It is easy to see that system (10) is T-periodic in t and, in view of ε being a small positive number and *Assumption* 2, is thus in the form ready for using averaging analysis [12]. Then, employing the averaging method gives:

$$\begin{cases} \dot{\hat{\bar{x}}}_i^{av} = -\varepsilon\bar{\phi}_i^{\,av}(\hat{\bar{x}}_i^{av}, \bar{x}_{-i}^g, \lambda) + O(\varepsilon^2)\mathbf{1} \\ \bar{\phi}_i^{\,av} = \frac{\bar{w}}{2\pi}\int_0^{\frac{2\pi}{\bar{w}}} \phi_i(\hat{\bar{x}}_i^{av} + a\cdot\sin(w_i t), \bar{x}_{-i}^g, \lambda)\sin(w_i t)\mathrm{d}t. \end{cases} \tag{13}$$

where $\mathbf{1}$ denotes the all-ones column vector with proper dimension. Using Taylor series expansion gives

$$\begin{aligned} \bar{\phi}_i^{\,av} &= \frac{a\bar{w}}{2\pi}\int_0^{\frac{2\pi}{\bar{w}}} \sin(w_i t)[\sin(w_i t)]^T \nabla_{\hat{\bar{x}}_i^{av}}\bar{\phi}_i \mathrm{d}t + O(a^2)\mathbf{1} \\ &= a[\hat{\delta}_{jk}]\nabla_{\hat{\bar{x}}_i^{av}}\bar{\phi}_i(\hat{\bar{x}}_i^{av}, \hat{\bar{x}}_{-i}^g + a\cdot\sin(w_i t), \lambda) + O(a^2)\mathbf{1} \\ &= a[\hat{\delta}_{jk}]\nabla_{\hat{\bar{x}}_i^{av}}\bar{\phi}_i(\hat{\bar{x}}_i^{av}, \hat{\bar{x}}_{-i}^g), \lambda) + O(a^2)\mathbf{1}, \end{aligned} \tag{14}$$

where

$$\hat{\delta}_{jk} = \frac{\bar{w}}{2\pi}\int_0^{\frac{2\pi}{\bar{w}}} \sin(w_i t)[\sin(w_i t)]^T \mathrm{d}t \tag{15}$$

Noting that $\frac{w_{ij}-w_{ik}}{\bar{w}}$ is an integer, integrating (15) gives

$$\hat{\delta}_{jk} = \begin{cases} 0, & w_{ij} \neq w_{ik}, \\ \frac{1}{2}, & w_{ij} = w_{ik}. \end{cases} \tag{16}$$

By substituting (16) into (14) and, in turn, (13), we get

$$\dot{\hat{\bar{x}}}_i^{av} = -\frac{\varepsilon a}{2}\nabla_{\hat{\bar{x}}_i^{av}}\bar{\phi}_i(\hat{\bar{x}}_i^{av}, \hat{\bar{x}}_{-i}^g, \lambda) + O(\varepsilon a^2 + \varepsilon^2)\mathbf{1}.$$

■

B. Stability Analysis of D-ESC Scheme

In this section, we will investigate the convergence property of D-ESC scheme. It is of convex-concave problems as presented in [17, 21, 9].

Theorem 1: Consider the D-ESC scheme as described in (11). Assume that *Assumption 2* holds. Then, the dynamic system (11) is SPAS on $[\varepsilon, a]$.

Proof: Consider the overall D-ESC scheme as described in (11). Let

$$\begin{aligned} V(\hat{\bar{x}}, \lambda) &= \frac{1}{2}\left\|\hat{\bar{x}} - \hat{\bar{x}}^*\right\|_{\Gamma_{\hat{\bar{x}}}} + \frac{1}{2}\left\|\lambda - \lambda^*\right\|_{\Gamma_\lambda^{-1}} \\ &= \sum_{\forall i\in\mathcal{I}} \frac{1}{2}\left\|\hat{\bar{x}}_i - \hat{\bar{x}}_i^*\right\|_{\Gamma_{\hat{\bar{x}}_i}} + \frac{1}{2}\left\|\lambda - \lambda^*\right\|_{\Gamma_\lambda^{-1}} \end{aligned}$$

be the Lyapunov function of the system, where $\hat{\bar{x}}^* \in \mathcal{C}$ and λ^* is the optimal point and certain dual variable attaining the minimum of the dual problem. Then, with the property (9) and let $\eta_i = (\hat{\bar{x}}_i - \hat{\bar{x}}_i^*)$, we have

$$\begin{aligned} \dot{V} &= \sum_{\forall i\in\mathcal{I}} (\hat{\bar{x}}_i - \hat{\bar{x}}_i^*)^T\Gamma_{\bar{x}_i}^{-1}\dot{\hat{\bar{x}}}_i + (\lambda-\lambda^*)\Gamma_\lambda^{-1}\dot{\lambda} \\ &= \sum_{\forall i\in\mathcal{I}} \eta^T\Gamma_{\bar{x}_i}^{-1}P_{\mathcal{C}_i}[-\varepsilon\bar{\phi}_i\cdot\sin(w_i t)] + (\lambda-\lambda^*)\Gamma_\lambda^{-1}\dot{\lambda} \\ &\leq \sum_{\forall i\in\mathcal{I}} \eta^T\Gamma_{\hat{\bar{x}}_i}^{-1}[-\varepsilon\bar{\phi}_i\cdot\sin(w_i t)] + (\lambda-\lambda^*)\Gamma_\lambda^{-1}\dot{\lambda} \end{aligned}$$

It follows from the above that the system is guaranteed to be stable as long as the corresponding system without projection is stable. In this regard, instead of directly analyzing the stability of the reduced system with projection, we can investigate the stability of the averaged version of the one without projection since it is in the form ready for using averaging analysis [12, Ch. 10]. Then, by *lemma* 1, we have the corresponding averaged system as follows

$$\dot{\hat{\bar{x}}}_i = -\frac{\varepsilon a}{2}\nabla_{\hat{\bar{x}}_i}\bar{\phi}_i(\hat{\bar{x}}_i, \hat{\bar{x}}_{-i}^g, \lambda) + O(\varepsilon a^2 + \varepsilon^2)\mathbf{1}.$$

Likewise, we can construct the Lyapunov function as above for the averaged system without considering the averaging error. Then, with (6) and the convexity of Φ we have

$$\begin{aligned} \dot{V} &= \sum_{\forall i\in\mathcal{I}} (\hat{\bar{x}}_i - \hat{\bar{x}}_i^*)^T\Gamma_{\bar{x}_i}^{-1}\dot{\hat{\bar{x}}}_i + (\lambda-\lambda^*)^T\Gamma_\lambda^{-1}\dot{\lambda} \\ &= \sum_{\forall i\in\mathcal{I}} \eta^T\nabla_{\hat{\bar{x}}_i}\bar{\phi}_i(\hat{\bar{x}}, \lambda) + (\lambda-\lambda^*)^T\Gamma_\lambda^{-1}\nabla_{\hat{\bar{x}}_i}\bar{\Phi}_i(\hat{\bar{x}}, \lambda) \\ &\leq [\Phi(\hat{\bar{x}}^*, \lambda) - \Phi(\hat{\bar{x}}, \lambda)] + [\Phi(\hat{\bar{x}}, \lambda) - \Phi(\hat{\bar{x}}, \lambda^*)] \\ &= [\Phi(\hat{\bar{x}}^*, \lambda) - \Phi(\hat{\bar{x}}^*, \lambda^*)] + [\Phi(\hat{\bar{x}}^*, \lambda^*) - \Phi(\hat{\bar{x}}, \lambda^*)] \leq 0, \end{aligned}$$

with equality if and only if $\hat{\bar{x}} = \hat{\bar{x}}^*$ and $\lambda = \lambda^*$. Hence, by Lasalle's theorem [12, Th. 4.4], it follows that the system in the averaged form is globally asymptotically stable with respect to the optimal point $\hat{\bar{x}}^*$. Further, by [18, Corollary 1] as well as the definition of SPAS, we can conclude that the original system is SPAS on $[\varepsilon, a]$. It means that so long as the parameters ε and a is chosen small enough, the system will eventually converge to the neighborhood of the optimal solution of the primal problem.

■

C. Singular Perturbation Analysis for the Overall System

As the system equipped with D-ESC has two-time-scale property, it is suitable for us to analyze the stability of the overall system using singular perturbation theory [12].

Theorem 2: Consider the networked large-scale systems (1) equipped with the D-ESC Scheme (11). Assume that *Assumption 1 and 2* hold. Then, the overall system is SPAS on $[\varepsilon, a]$.

Proof: It can be followed from the ESC scheme as shown in Fig. 1 that the transient dynamics when achieving its equilibrium has an isolated root $x = [\hat{\bar{x}} + a\cdot\sin(wt), \bar{x}^r]$, where x is the state of the dynamic system (1) and $\bar{x}^r$ denotes the states that do not directly contribute the output and constraint. In addition, according to *Theorem* 1, the reduced system (11) is shown to be SPAS on $[\varepsilon, a]$. Further,

[7] Here, 'gcd' refers to the greatest common divisor.

with *Assumption* 1, it is straightforward to conclude that the boundary-layer system is also globally asymptotically stable in the domain $\mathcal{C}$, uniformly in $\forall[\hat{\tilde{x}}+a\cdot\sin(wt)]\in\mathcal{C}$. Then, by using [29, Lem 1], it follows that the overall system equipped with the D-ESC Scheme (11) is SPAS on $[\varepsilon, a]$.

■

VI. Conclusion and Future Work

In this paper, we have studied the distributed optimization problem of the steady-state performance of networked large-scale systems under constraints. In particular, we have proposed a new D-ESC approach to solve this problem. This approach employs a primal-dual method and ESC scheme which is utilized to do efficient gradient estimation. Thus, it can be implemented in a distributed way without knowing the specific cost functions as well as constraints so long as they can be measured, which is very important in practice in that most systems to be optimized are large-scale and subject to a great amount of uncertainty in its dynamics. However, issues regarding communication and computation, such as *uncertainty*, *asynchronism* and *varying topology* that is inherent to networked large-scale systems, are not taken into account in this paper and will be dealt with in future work with the help of consensus theory. In addition, the proposed approach is expected to be extended to deal with invex-concave and time-varying problems. Moreover, we are trying to apply it into real problems for energy efficiency.

References

[1] A. Aquino-Lugo. *Distributed and decentralized control of the power grid*. PhD thesis, University of Illinois at Urbana-Champaign, 2010.

[2] Kartik B. Ariyur and Miroslav Krstic. *Real-time optimization by Extremum-seeking Control*. Hoboken, NJ : Wiley Interscience, c2003., 2003.

[3] K. J. Arrow, L. Hurwicz, H. Uzawa, and H. B. Chenery. *Studies in linear and non-linear programming*. Stanford University Press, Stanford, 1958.

[4] J. Baillieul and P. J. Antsaklis. Control and communication challenges in networked real-time systems. In *Proceedings of the Ieee*, volume 95, pages 9–28, June 2007.

[5] T. Basar and G. J. Olsder. *Dynamic noncooperative game theory*. Society for Industrial Mathematics, 1999.

[6] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex analysis and optimization*. Athena Scientific, 2003.

[7] F. Borrelli and T. Keviczky. Distributed LQR design for identical dynamically decoupled systems. *IEEE Transactions on Automatic Control*, 53:1901–1912, 2008.

[8] W. B. Dunbar. Distributed receding horizon control of dynamically coupled nonlinear systems. *IEEE Transactions on Automatic Control*, 52:1249–1263, 2007.

[9] D. Feijer and F. Paganini. Stability of primaldual gradient dynamics and applications to network optimization. *Automatica*, 46:1974–1981, 2010.

[10] P. Frihauf, M. Krstic, and T. Basar. Nash equilibrium seeking for games with non-quadratic payoffs. In *Proceedings of 49th IEEE Conference on Decision and Control*, pages 881–886, December14–17 2010.

[11] T. Keviczky, F. Borrelli, and G. J. Balas. Decentralized receding horizon control for large scale dynamically decoupled systems. *Automatica*, 42:2105–2115, 2006.

[12] H. K. Khalil. *Nonlinear systems*. Prentice Hall, 2002.

[13] M. Krstic and H. H. Wang. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36:595–601, 2000.

[14] S. H. Low and D. E. Lapsley. Optimization flow control. I. Basic algorithm and convergence. *IEEE/ACM Transactions on Networking*, 7:861–874, 1999.

[15] Jason R Marden, S Ruben, and L Pao. A model-free approach to wind farm control using game theoretic methods. *IEEE Transactions on Control Systems Technology*, 2013.

[16] A. Nedic and A. Ozdaglar. Distributed subgradient methods for Multi-agent optimization. *IEEE Transactions on Automatic Control*, 54:48–61, 2009.

[17] A. Nedic and A. Ozdaglar. Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142:205–228, 2009.

[18] D. Nesic and A. R. Teel. Input-to-state stability for nonlinear time-varying systems via averaging. *Mathematics of Control, Signals, and Systems (MCSS)*, 14:257–280, 2001.

[19] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. In *Proceedings of the IEEE*, pages 215–233, Jan 2007.

[20] R. L. Raffard, C. J. Tomlin, and S. P. Boyd. Distributed optimization for cooperative agents: Application to formation flight. In *Proceedings of 43rd IEEE Conference on Decision and Control*, pages 2453–2459, December14–17 2004.

[21] A. Rantzer. Dynamic dual decomposition for distributed control. In *American Control Conference*, pages 884–888, June10–12 2009.

[22] J. B. Rawlings and B. T. Stewart. Coordinating multiple optimization-based controllers: New opportunities and challenges. *Journal of Process Control*, 18:839–845, 2008.

[23] M. A. Rotea. Analysis of multivariable extremum seeking algorithms. In *Proceedings of the 2000 American Control Conference*, volume 1, pages 433–437, June 2000.

[24] S. Samar, S. Boyd, and D. Gorinevsky. Distributed estimation via dual decomposition. In *Proceedings of the European Control Conference*, pages 1511–1516, July2–5 2007.

[25] E. Semsar-Kazerooni and K. Khorasani. Multi-agent team cooperation: A game theory approach. *Automatica*, 45:2205–2213, 2009.

[26] S. Skogestad. Self-optimizing control: The missing link between steady-state optimization and control. *Computers & Chemical Engineering*, 24:569–575, 2000.

[27] M. S. Stankovic, K. H. Johansson, and D. M. Stipanovic. Distributed seeking of Nash equilibria in mobile sensor networks. In *Proceedings of 49th IEEE Conference on Decision and Control*, pages 5598–5603, December14–17 2010.

[28] Y. Tan, W. Moase, C. Manzie, D. Neic, and I. Mareels. Extremum seeking from 1922 to 2010. In *Proceedings of 29th Chinese Control Conference*, pages 14–26, July29–31 2010.

[29] Y. Tan, D. Nesic, and I. Mareels. On non-local stability properties of extremum seeking control. *Automatica*, 42:889–903, 2006.

[30] Y. Tan, D. Nesic, I. M. Y. Mareels, and A. Astolfi. On global extremum seeking in the presence of local extrema. *Automatica*, 45:245–251, 2009.

[31] H. Terelius, U. Topcu, and R. M. Murray. Decentralized Multi-agent optimization via dual decomposition. In *Proceedings of the 18th IFAC World Congress*, pages 11245–11251, 2011.

[32] Y. Wakasa, M. Arakawa, K. Tanaka, and T. Akashi. Decentralized model predictive control via dual decomposition. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 381–386, December 9–11 2008.

[33] S. L. Waslander, G. Inalhan, and C. J. Tomlin. Decentralized optimization via Nash bargaining. In *Theory and Algorithms for Cooperative Systems*, pages 565–583. World Scientific Pub. Co. Inc., Destin, 2003.

[34] L. Xiao, M. Johansson, and S. P. Boyd. Simultaneous routing and resource allocation via dual decomposition. *IEEE Transactions on Communications*, 52:1136–1144, 2004.

[35] D. Yuan, S. Xu, and H. Zhao. Distributed Primal-Dual subgradient method for Multi-agent optimization via consensus algorithms. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 41:1715–1724, 2011.

[36] M. Zhu and S. Martinez. On distributed convex optimization under inequality and equality constraints. *IEEE Transactions on Automatic Control*, 57(1):151–164, 2012.