

# Contraction and Robustness of Continuous Time Primal-Dual Dynamics

Hung D. Nguyen<sup>ID</sup>, Thanh Long Vu<sup>ID</sup>, Konstantin Turitsyn<sup>ID</sup>, and Jean-Jacques Slotine<sup>ID</sup>

**Abstract**—The Primal-dual (PD) algorithm is widely used in convex optimization to determine saddle points. While the stability of the PD algorithm can be easily guaranteed, strict contraction is nontrivial to establish in most cases. This letter focuses on continuous, possibly non-autonomous PD dynamics arising in a network context, in distributed optimization, or in systems with multiple time-scales. We show that the PD algorithm is indeed strictly contracting in specific metrics and analyze its robustness establishing stability and performance guarantees for different approximate PD systems. We derive estimates for the performance of multiple time-scale multi-layer optimization systems, and illustrate our results on a PD representation of the Automatic Generation Control of power systems.

**Index Terms**—Primal-dual dynamics, continuous optimization, strict contraction, robustness, hierarchical architecture.

## I. INTRODUCTION

MULTI-SCALE optimization is ubiquitous in both natural and artificial systems. Multiple time-scales have long been viewed as a fundamental organizing principle in the modular architecture and evolution of complex systems [1]. In engineering, such layering provides a powerful design architecture for decomposing multiple stage decision processes in networked infrastructure [2]. In cyber-physical networks such as the power grid, the structure of the dynamic equations is strongly constrained by physical laws. However, as a number of recent works have shown, e.g., [3]–[5], the nature of these equations admits a natural optimization based perspective. To a large extent, the underlying ideas behind the approaches exploited in those works can be traced back to the classical saddle point problem, and its associated primal-dual (PD) algorithm.

Manuscript received March 6, 2018; revised May 23, 2018; accepted June 9, 2018. Date of publication June 14, 2018; date of current version July 6, 2018. The Work of H. D. Nguyen was supported by in part by NTU SUG, in part by Siebel Fellowship, and in part by VEF. The Work of K. Turitsyn was supported by NSF CAREER under Award 1554171 and Award 1550015. Recommended by Senior Editor J. Daafouz. (Corresponding author: Hung D. Nguyen.)

H. D. Nguyen is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: hungtd@mit.edu).

T. L. Vu, K. Turitsyn, and J.-J. Slotine are with the Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Digital Object Identifier 10.1109/LCSYS.2018.2847408

The saddle-point problem, first considered in the context of market equilibrium in economics, appears when the system is simultaneously minimizing a function over one set of its variables and maximizing it over the other variable set. Due to the unique characteristic of the cost function, namely, being convex in the first set variables while being concave in the second, the primal-dual algorithm consists of implementing gradient descent compatibly with the convexity/concavity of the cost function. This algorithm was introduced in the classic work of Arrow *et al.* [6]. Recently, the body of literature on this algorithm has been rapidly growing due to its efficiency in decentralized optimization in network applications [7]–[10]. Similar saddle point problems also appear naturally in the context of machine learning, e.g., in support vector machine representations [11] and in the adversarial training of deep networks [12].

## A. The PD Algorithm in a Distributed Context

In its simplest form, the primal dual (PD) dynamics minimizes the function  $g(x) \in \mathbb{R}^n \rightarrow \mathbb{R}$  subject to a set of linear or nonlinear constraints  $h(x) = 0$  with  $h(x) \in \mathbb{R}^m \rightarrow \mathbb{R}^m$ . The corresponding Lagrangian function is given by

$$\ell(x, v) = g(x) + v^T h(x), \quad (1)$$

with  $v \in \mathbb{R}^m$  being the dual variables. We denote the full state of the system by  $z = (x, v) \in \mathbb{R}^{n+m}$ . The continuous time primal dual algorithm defines a dynamic system in  $x(t)$  and  $v(t)$  described by

$$\dot{x} = -\partial_x \ell = -\partial_x g - (\partial_x h)^T v \quad (2)$$

$$\dot{v} = +\partial_v \ell = h \quad (3)$$

There are two common applications of primal-dual dynamics most frequently discussed in the literature. First, it can be naturally applied to design distributed continuous time optimization systems using the Lagrangian relaxation type of approach. Within this formulation, the large-scale optimization problem is represented in a superposition form  $g(x) = \sum_k g_k(\mathcal{X}_k)$  with  $\mathcal{X}_k$  denoting the set of variables for subsystem  $k$ . Coupling between the subsystems is enforced by constraining subsets of variables in different subproblems to be equal to each other. These additional constraints are represented as  $h(x) = E^T x$ , with  $E$  being an incidence matrix of a directed graph defining equivalencies between variable replicas. In this setting, every subproblem is coupled to common dual variables but not other subproblem variables directly.

Every subproblem can then be solved by an individual agent, while the dual dynamics gradually adjusts the dual variables until the equivalence constraints are satisfied.

Another typical application arises in network flows of natural or artificial nature characterized by some conservation laws. Assuming that some subset of variables represents the flows of the conserved quantities, the conservation laws are expressed as  $h(x) = Ex - q$ , with  $E$  being the incidence matrix reflecting the topology of interconnection of individual subsystems, and  $q$  the vector of external source/sink flows. In equilibrium, the total flow on every node of the interconnection is balanced. However, during the transient dynamics violation can occur due to non-zero  $\dot{v}$ . These violations may be interpreted as the accumulation of the transported quantity on the node storage elements. In traditional circuits, the  $\dot{v}$  terms represent charging of effective node capacitance. Whenever such a formulation of physical equations exists, one can naturally solve the distributed optimization problem complementing the Lagrangian with additional terms representing the objective of the controller.

Unless otherwise stated here we assume the constraints of the form  $h(x) = Ex - q$ , which leads to the following formulation

$$\dot{x} = -\partial_x g - E^T v \quad (4a)$$

$$\dot{v} = Ex - q. \quad (4b)$$

## B. Contributions

In Section II, we establish the strict contraction of continuous PD dynamics in the form given by (4), by constructing explicitly contracting metrics and estimating the corresponding contraction rates. Next, in Section III, we derive several new results concerning the robustness of the PD systems including the bounds on the long-term steady-state errors induced by various types of disturbances, inaccurate estimations, and multiple time-scale optimization. We dedicate Section IV to present a relevant power systems example where we present the AGC problem in a PD form then illustrate the error bounds derived in Section III. Note that our results can apply to non-autonomous PD dynamics with general objective functions, while most recent related work deals with autonomous PD dynamics [13] or quadratic objective functions [14]. The ISS analysis in [15], which characterizes error bounds to fixed saddle points, is relevant to the estimates derived in this letter. However, we quantify errors relative to the time-varying instantaneous optimum (Lemma 2), as well as those relative to the time-varying trajectory induced by imperfect measurements (Lemma 3).

## II. STRICT CONTRACTION OF PD DYNAMICS

Throughout this letter, we use  $\|\cdot\|$  to denote the norm in which the considered system is strictly contracting. Similarly,  $\mu(A)$  denotes the matrix measure of  $A$  corresponding to the discussed norm. In particular, for the 2-norm, one has  $\mu(A) = \lambda_{\max}(A + A^T)/2$  where  $\lambda_{\max}$  is the maximal eigenvalue.

We proceed by reviewing the basics of contraction analysis for nonlinear dynamical systems. For holistic descriptions on this topic, see [16], [17]. Let us consider a nonlinear dynamical

system  $\dot{z} = f(z, t)$  where  $f$  is a continuous and sufficiently smooth function of the state variable  $z$ . The infinitesimal dynamics can be given as  $\delta\dot{z} = \partial_z f \delta z$ . Contraction analysis can characterize the dynamics of the distance between two close trajectories in some weighted two norm defined as:  $\|\delta z\| = \|\delta y\|_2$  where we introduce a differential coordination transformation  $\delta y = \Theta \delta z$  with an invertible metric transformation  $\Theta$ . The rate of change of the distance can be calculated accordingly  $d\|\delta z\|_2^2/dt = 2\delta y^T F \delta y$  with  $F = \dot{\Theta}\Theta^{-1} + \Theta\partial_z f\Theta^{-1}$  being the generalized Jacobian. Whenever the symmetric part of the generalized Jacobian is uniformly negative definite, i.e., there exists  $\beta > 0$  s.t.  $\mu(F) \leq -\beta < 0$ , the system is (strictly)  $\beta$ -contracting and all trajectories will converge exponentially towards each other with a contraction rate larger or equal to  $\beta$ .

The basic contraction property of the primal dual dynamics with respect to the identity metric is simple to establish [17]. Specifically, consider a virtual displacement between the two close trajectories characterized by the vector  $\delta z = (\delta x, \delta v)$ . These displacement vectors are described by:

$$\delta\dot{x} = -H\delta x - E^T\delta v \quad (5a)$$

$$\delta\dot{v} = E\delta x \quad (5b)$$

with  $H = \partial_{xx}f$  being the Hessian of the objective function. We further assume that the objective function is strictly convex in  $x$ , thus its corresponding Hessian is positive definite  $H > 0$ . One can easily see that the primal-dual dynamics described by (4) is contracting with respect to the traditional Euclidean norm:  $\|\delta z\|_2^2 = \delta x^T \delta x + \delta v^T \delta v$ :

$$\frac{d}{dt}\|\delta z\|_2^2 = -2\delta x^T H \delta x \leq 0 \quad (6)$$

Among other things, this result implies that the distance  $r(z(t), z^*) = \|z(t) - z^*\|$  between the current point and any equilibrium satisfying the KKT conditions is a non-increasing function, a well-known result since the classical works on PD dynamics [6]. Moreover, it establishes the connection between this original result and more recent approach to PD systems via Krasovskii type Lyapunov functions [7]. Indeed, the existence of a Krasovskii Lyapunov function implies contraction of the system, which in turn verifies that the distance between any point and any equilibrium only decreases.

However, this system is not strictly contracting; there may be close trajectories that don't eventually get closer to each other in Euclidean metrics. This is indeed the case for trajectories with the same  $x$  but different  $v$ . Moreover, whenever the matrix  $E$  is not full row rank, the optimum of the original system may not be unique, and the system may converge to different equilibria. In this case, there is no strict contraction because the distance between two trajectories starting from different equilibria does not change. However, for systems with full rank  $E$  the question arises, whether strict contraction of the trajectories can be established in some other metric.

### A. Strict Contraction of PD Dynamics

In the following we shall develop a metric which certifies strict contraction of PD systems. Consider a coordinate

transformation  $\delta y = \Theta \delta z$  with invertible “skew” metric transformation  $\Theta$ :

$$\Theta = \begin{bmatrix} I & \alpha E^T \\ 0 & (I - \alpha^2 EE^T)^{\frac{1}{2}} \end{bmatrix} \quad (7)$$

where  $I$  denotes the identity matrix of appropriate dimensions. Then, for the two neighbour trajectories with the virtual displacement  $\delta z$ , we introduce the distance  $\|\delta z\|_\alpha^2 = \delta y^T \delta y = \delta z^T \Theta^T \Theta \delta z$ . Following the arguments from the previous sections we arrive at the generalized Jacobian  $F = -\Theta^{-T} Q \Theta^{-1}$  with

$$Q = \begin{bmatrix} H - \alpha E^T E & \frac{\alpha}{2} EH \\ \frac{\alpha}{2} HE^T & \alpha EE^T \end{bmatrix} \quad (8)$$

Theorem below establishes sufficient conditions for strict contraction of the primal-dual algorithm in this metric.

*Lemma 1:* If  $\alpha$  satisfies

$$0 < \alpha < \frac{1}{\max\{\|E\|_2, \|EH^{-1/2}\|_2^2 + \|H\|_2/4\}} \quad (9)$$

the primal-dual system defined in (4) is strictly contracting in metric  $\Theta$ , with rate  $\beta = \mu_2(-\Theta^{-T} Q \Theta^{-1})$ .

*Proof:* The condition  $\alpha < 1/\|E\|$  ensures not only that the metric transformation  $\Theta$  is invertible but also the distance  $\|\delta z\|_\alpha^2$  is a positive definite form:

$$\begin{aligned} \|\delta z\|_\alpha^2 &= \delta z^T \Theta^T \Theta \delta z \\ &= \|\delta x + \alpha E^T \delta v\|_2^2 + \delta v^T (I - \alpha^2 EE^T) \delta v. \end{aligned} \quad (10)$$

The rate of change of this form, according to (4) is given by

$$\begin{aligned} \frac{d}{dt} \|\delta z\|_\alpha^2 &= -2\delta x^T (H - \alpha E^T E) \delta x - 2\alpha \delta v^T EH \delta x - 2\alpha \delta v^T EE^T \delta v \\ &= -2\delta x^T (H - \alpha E^T E - \frac{\alpha}{4} H^2) \delta x - 2\alpha \|E^T \delta v\|_2^2 + \frac{1}{2} H \delta x \|_2^2 \end{aligned}$$

Matrix  $H - \alpha E^T E - \frac{\alpha}{4} H^2$  is positive definite whenever  $\alpha$  satisfies (9), and the system is therefore strictly contracting. For the considered system, the rate of change of the distance  $\|\delta z\|_\alpha^2 = \|\Theta \delta z\|_2^2$  is given by  $\frac{d}{dt} \|\delta z\|_\alpha^2 = -2\delta z^T Q \delta z$ ,  $Q > 0$ . The guaranteed contraction rate is given by  $\beta = \lambda_{\min}(\Theta^{-T} Q \Theta^{-1})$  which is easily proved by noting that  $-\delta z^T Q \delta z = -(\Theta \delta z)^T (\Theta^{-T} Q \Theta^{-1}) (\Theta \delta z) \leq -\lambda_{\min}(\Theta^{-T} Q \Theta^{-1}) \|\Theta \delta z\|_2^2$ . ■

Note that, for linear and autonomous systems, the metric  $\Theta^T \Theta$  above is equivalent to the strict quadratic Lyapunov functions for PD dynamics presented in [14]. However, the main advantage of our approach stems from the fact that contraction analysis applies to a more general class of systems which can be non-autonomous. Many important classes of problems are non-autonomous in nature including the optimization framework considered in this letter.

### III. APPLICATIONS

While the PD dynamics is an extremely flexible framework applicable to a broad variety of continuous time optimization problems, its perfect realization is not viable in most of the practical situations. In this section, we analyze stability and performance of quasi-PD systems that approximate the “true”

PD dynamics. Throughout the section, we adopt a number of assumptions and definitions reviewed below. We consider a system characterized by the Lagrangian  $\ell(x, v, t)$  of the form (1) and the “true” PD dynamics expressed compactly as

$$\dot{z}^{\text{pd}} = f(z^{\text{pd}}, t) \quad (11)$$

where  $z^{\text{pd}} = (x^{\text{pd}}, v^{\text{pd}})$ . We assume that the system is contracting with respect to the uniform metric associated with the norm  $\|z\| = \|\Theta z\|$  with some constant matrix  $\Theta$  as presented in Section II. We assume that the system is strictly  $\beta$ -contracting.

While many of the results can be easily extended to a more general case of non-uniform metric  $\Theta$  explicitly depending on  $z$ , for the sake of simplicity we restrict the discussion only to the uniform case. In the following, we show that the contraction rate with respect to such a metric can be naturally used to characterize the performance of more realistic approximately primal-dual systems.

In many practical situations, the continuous time PD algorithm is utilized in a non-stationary setting where the system is subject to constantly changing external conditions. In this case, the PD dynamics allows the system to track the optimal operating condition which also changes in time. For example, in power system context, the secondary frequency control can keep the system close to the optimal power flow solutions as the external factors such as the load power consumption or renewable generation change. Typically there is a time-scale separation between the fast PD dynamics and slow changes of external parameters. In this case, the deviations from the optimum are small enough and can be safely ignored. However, in a more general context establishing rigorous bounds on the deviations from the optimum is essential for certifying the safety and performance of the systems.

In the following, we assume, that the Lagrangian  $\ell(x, v, t)$  is explicitly dependent on time, and characterize this dependence implicitly through the position of the instantaneous optimum  $z^*(t) = (x^*(t), v^*(t))$ . Also, we explicitly assume that the rate of change  $\|\dot{z}^*\|$  of such instantaneous optimum is bounded. Then strict contraction of the PD dynamics provides us with a natural for quantification of the deviations from the optimum.

*Lemma 2:* Consider now the distance  $r(t) = \|z^{\text{pd}}(t) - z^*(t)\|$  between the state  $z^{\text{pd}}(t)$  and the instant optimum  $z^*(t)$ . This distance satisfies the following differential inequality:

$$\dot{r} \leq -\beta r + \|\dot{z}^*\| \quad (12)$$

*Proof:* This result is proven by direct differentiation of  $r^2(t)$  and application of the contraction property. The term  $\beta r$  represents the contraction of the fixed equilibrium system, while the term  $\|\dot{z}^*\|$  accounts for the motion of the instantaneous equilibrium point  $z^*(t)$ . ■

*Corollary 1:* In the steady state  $t \rightarrow \infty$ , whenever the rate of equilibrium point movement is bounded the “true” PD system above is confined to the ball

$$\|z^{\text{pd}}(t) - z^*(t)\| \leq \frac{1}{\beta} \sup_t \|\dot{z}^*\|. \quad (13)$$

#### A. Robustness of PD Systems

Next, we consider the systems that deviate from the “true” PD dynamics. Our primary motivation is the multi-scale

optimization system where the decisions are made by different layers on multiple time-scales (see [18], [19]). It is common in this setting for the higher layers of the hierarchy to have limited observability of the lower layer states. Most commonly, the optimization logic on higher layers either relies on imperfect observations or assumes that the faster lower layers have already equilibrated. To model this setting, we represent the approximate primal-dual dynamics by

$$\dot{z} = f(z, t) + d(z, t) \quad (14)$$

where  $d(z, t) = f(\hat{z}, t) - f(z, t)$  represents the substitution of true signal  $z = (z_o, z_u)$  with its estimate  $\hat{z} = (z_o, \hat{z}_u)$ . In this section we assume that the function  $f$  is Lipschitz, so that the following two inequalities hold:

$$\|d(z, t)\| = \|f(\hat{z}, t) - f(z, t)\| \leq \xi \|\hat{z} - z\| \quad (15)$$

$$\|f(z, t)\| = \|f(z, t) - f(z^*, t)\| \leq \eta \|z - z^*\| \quad (16)$$

We now prove the following intermediate lemma (see also [20] and well as [19]).

*Lemma 3:* In steady-state, the distance  $r^{\text{pd}}(t) = \|z(t) - z^{\text{pd}}(t)\|$  between the approximate and true primal-dual dynamics systems satisfies the differential inequality:

$$\dot{r}^{\text{pd}} \leq -\beta r^{\text{pd}} + \|d\| \quad (17)$$

*Proof:* Consider a trajectory  $\tilde{z}(t)$  following the original flow, i.e., satisfying  $\dot{\tilde{z}} = f(\tilde{z}, t)$  and starting at time  $t$  at  $z(t)$ , i.e.,  $\tilde{z}(t) = z(t)$ . Trajectory  $\tilde{z}(t)$  together with  $z^{\text{pd}}(t)$  can be considered as two individual trajectories of the same system which is contracting at the rate of  $\beta$ . Strict contraction of the original system implies that the distance between these trajectories will decrease as:

$$\|\tilde{z}(t + dt) - z^{\text{pd}}(t + dt)\| \leq (1 - \beta dt) \|z(t) - z^{\text{pd}}(t)\| \quad (18)$$

where we have utilized the fact that  $z(t) = \tilde{z}(t)$  at time  $t$ . At the same time, the distance between  $\tilde{z}(t)$  and  $z(t)$  in the same interval has increased by at most  $dt \|d(z(t), t)\|$ . Therefore, we have that

$$\|\tilde{z}(t + dt) - z(t + dt)\| \leq \|d(z(t), t)\| dt. \quad (19)$$

Combining (18) and (19) via triangle inequality:  $\|\tilde{z}(t + dt) - z(t + dt)\| - \|z(t + dt) - z^{\text{pd}}(t + dt)\| \leq \|\tilde{z}(t + dt) - z^{\text{pd}}(t + dt)\|$ , and taking the limit  $dt \rightarrow 0$ , yields (17). ■

*Corollary 2:* After exponential transients at rate  $\beta > 0$ , the distance between the non-ideal and ideal PD can be bounded as

$$\|z(t) - z^{\text{pd}}(t)\| \leq \frac{\xi}{\beta} \sup_t \|\hat{z}(t) - z(t)\| \quad (20)$$

*Proof:* From Lemma 3 one has the following when the system settles down:  $\|z(t) - z^{\text{pd}}(t)\| \leq \sup_t \|d(z(t), t)\| / \beta$ . Combining this relation and the inequality (15), yields (20). ■

Next, we consider a setting where the actual system does not follow exactly the primal-dual dynamics, although the PD system is a reasonable approximation. In practice, this situation can occur for a variety of reasons, for example, due to fast degrees of freedom in plant dynamics lacking the PD structure, or due to imperfect observers introducing additional delays in

the system. While both settings can be modeled in the same way, for exposition purposes we restrict the discussion only to the case of imperfect observers. We assume, that the subset of directly unobservable variables  $z_u$  is estimated by a separate observer system  $\hat{z}_u = f_u(\hat{z}_u, z_u)$  that satisfies the following conditions:

- 1) A subset of the observer states,  $\hat{z}_u$  is an asymptotically unbiased estimate of  $z_u$ , so that for constant  $z_u$ , the observer converges to the manifold  $f_u(\hat{z}_u, z_u) = 0$  satisfying  $\hat{z}_u = z_u$ .
- 2) Dynamics of the observer is partially contracting with respect to  $\hat{z}_u$  with a contraction rate of  $\hat{\beta}$ .

The following formal results allow us to characterize the performance of the system.

*Lemma 4:* Whenever  $\hat{\beta} > \xi$  and  $\|z - z^*\|$  is upper bounded, the long-term steady state error of the observer is bounded by

$$\|\hat{z} - z\| \leq \frac{\eta}{\hat{\beta} - \xi} \sup_t \|z - z^*\| \quad (21)$$

*Proof:* Applying Lemma 2 and the Lipschitz bounds we obtain

$$\|\hat{z} - z\| \leq \frac{1}{\hat{\beta}} \sup_t \|\dot{z}_u\| \quad (22a)$$

$$\leq \frac{1}{\hat{\beta}} \sup_t (\|f(z, t)\| + \|d(t)\|) \quad (22b)$$

$$\leq \frac{1}{\hat{\beta}} \sup_t (\eta \|z - z^*\| + \xi \|\hat{z} - z\|) \quad (22c)$$

where we have used the relation  $\|\dot{z}_u\| \leq \|\dot{z}\| \leq \|f(z, t)\| + \|d(t)\|$  to arrive at (22b) from (22a). Solving for  $\sup_t \|\hat{z} - z\|$  yields (21). ■

*Theorem 1:* Given the conditions stated in Lemma 4 and  $\beta(\hat{\beta} - \xi) - \eta\xi > 0$ , the steady state optimal tracking error is upper bounded

$$\|z - z^*\| \leq \frac{\hat{\beta} - \xi}{\beta(\hat{\beta} - \xi) - \eta\xi} \sup_t \|\dot{z}^*\| \quad (23)$$

*Proof:*

$$\|z - z^*\| \leq \|z - z^{\text{pd}}\| + \|z^{\text{pd}} - z^*\| \quad (24a)$$

$$\leq \frac{1}{\beta} \sup_t (\xi \|\hat{z} - z\| + \|\dot{z}^*\|) \quad (24b)$$

$$\leq \frac{1}{\beta} \sup_t \left( \frac{\eta\xi}{\hat{\beta} - \xi} \|z - z^*\| + \|\dot{z}^*\| \right) \quad (24c)$$

where we use relation (21) to arrive at the last inequality from (24b). By solving for  $\sup_t \|z - z^*\|$  we obtain (23). ■

Note, that in the limit of perfect observer with  $\hat{\beta} \rightarrow \infty$  one recovers the bound (13).

*Corollary 3:* The distance between the perturbed and “true” primal-dual dynamics can be bounded as

$$\|z - z^{\text{pd}}\| \leq \frac{1}{\beta} \frac{\eta\xi}{\beta(\hat{\beta} - \xi) - \eta\xi} \sup_t \|\dot{z}^*\| \quad (25)$$

Corollary 3 can be easily proved by combining the results from Corollary 2 and Lemma 4.

The bounds derived in this section are illustrated with a practical power system in Section IV. While their practical relevance should be assessed in the context of specific systems

with well-defined structural features, the bounds are tight if one assumes that only Lipschitz constants are known. For example, tightness of the bound (13) can be established by considering the simple scalar ODE  $\dot{z}(t) = -z(t) + t$  with  $z(0) = 0$ .

### B. Performance of Layered Optimization Architectures

Our results could be naturally applied to multi-layer optimization systems commonly occurring in nature and engineering. In these systems, each layer typically performs its own optimization [1], [2], [21], and interacts with other layers. Usually, the dynamics of the layers are separated in time-scales, so that the dynamics of higher levels is slower in comparison to that of lower ones. In engineering systems, the algorithms employed on the individual layers are often designed with two assumptions in mind: that the lower layer has converged to its optimal equilibrium, and that the higher layer inputs can be assumed to be constant. In this section, we study a broader class of such systems, constrained only to be Lipschitz and contracting.

Mathematically, the setting above can be expressed by introducing the subsets of PD variables  $z_k$  corresponding to different layers of optimization. For notational simplicity, we assume that each layer interacts directly with two neighboring layers and the ‘‘true’’ PD dynamics is described by

$$\dot{z}_k = f_k(z_{k-1}, z_k, z_{k+1}, t) \quad (26)$$

From the viewpoint of layer  $k$ , the function  $z_{k-1}$  can be considered as an exogenous factor, which affects the position of instant equilibrium  $z_k^* = z_k^*(z_{k-1}(t), t)$ . On the other hand, we assume that each layer  $k$  is designed under the assumption that all the faster layers have equilibrated and so  $z_{k+1}$  is replaced by  $z_{k+1}^*(z_k, t)$ . In this case, following the same logic as in previous section, the actual dynamics can be represented as  $\dot{z}_k = f_k + d_k$  with

$$d_k = f_k(z_{k-1}, z_k, z_{k+1}^*(z_k, t), t) - f_k(z_{k-1}, z_k, z_{k+1}, t) \quad (27)$$

*Theorem 2:* Consider a multi-layer optimization system described above. Assume that  $f_k$  is Lipschitz with respect to  $z_k$  and  $z_{k+1}$  with constants denoted as  $\eta_k, \xi_k$  respectively. Furthermore, assume that each function  $z_k^*(z_{k-1}, t)$  is also Lipschitz with respect to  $z_{k-1}$ , with Lipschitz constant  $\rho_k$ .

Let  $\gamma_k = 1 - \eta_k \tau_{k+1} \rho_{k+1}$ . For small enough Lipschitz constants, such that  $\gamma_k > 0$  and  $\gamma_k \beta_k > \xi_k$ , the system is stable and its performance is characterized by inequalities  $\|z_k - z_k^*\| \leq \tau_k \sup_t \|\dot{z}_k^*\|$  with  $\tau_k$  given by (13) or (23) for the lower layer, and for the higher ones by

$$\tau_k = \frac{\gamma_k}{\gamma_k \beta_k - \xi_k}. \quad (28)$$

*Proof:* We start by bounding the term  $d_k$ . Using the definition (27) we have

$$\begin{aligned} \|d_k\| &\leq \eta_k \|z_{k+1} - z_{k+1}^*\| \\ &\leq \eta_k \tau_{k+1} \sup_t \|\dot{z}_{k+1}^*\| \\ &\leq \eta_k \tau_{k+1} \rho_{k+1} \|\dot{z}_k\| \\ &\leq \eta_k \tau_{k+1} \rho_{k+1} \sup_t (\xi_k \|z_k - z_k^*\| + \|d_k\|) \end{aligned}$$

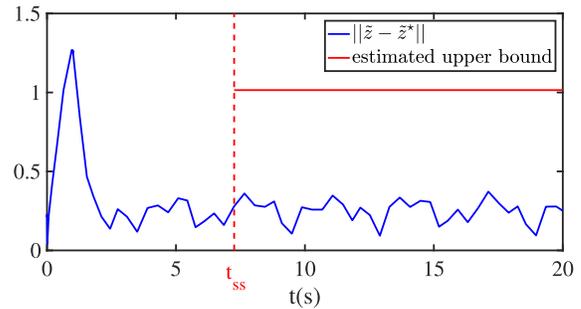


Fig. 1. The long-term steady-state error is upper curved by a bound (the red line) proportional to instantaneous optimum rate in Theorem 1.

Hence, whenever  $\eta_k \tau_{k+1} \rho_{k+1} < 1$  we have

$$\|d_k\| \leq \frac{\xi_k \|z_k - z_k^*\|}{1 - \eta_k \tau_{k+1} \rho_{k+1}} \quad (29)$$

Next, as long as the system  $f_k$  is contracting, we have  $\|z_k - z_k^*\| \leq (\|\dot{z}_k^*\| + \|d_k\|)/\beta_k$ . Using (29), assuming  $\xi_k < \beta_k(1 - \eta_k \tau_{k+1} \rho_{k+1})$ , we arrive at (28). ■

Note, that while multi-scale PD dynamics have been a motivation for this section analysis, the results apply more broadly to general multi-scale contracting systems, not necessarily optimization ones. On the other hand, many practical iterative optimization algorithms in discrete time can be viewed as perturbed versions of the baseline continuous dynamics and thus admit analysis with the proposed techniques. Our multiple time-scale results are also applicable to the *virtual* contracting systems used to analyze synchronization phenomena [22]. In particular, quorum sensing strategies [23], [24] can be used to coordinate multiple dynamics.

## IV. NUMERICAL SIMULATIONS

In this section, we illustrate the applications of the derived bounds by considering the so-called Automatic Generation Control (AGC) of power systems [25], [26]. There has been a lot of interest recently in exploring the optimization perspective on the frequency control, see [3], [8] for review. As a proof of concept, here we look at a simplified AGC model designed to restore the frequency of the system.

The simplified rescaled AGC model can be represented in PD form, with the Lagrangian

$$\ell = \frac{1}{2} \omega^T D \omega + \omega^T (B^{1/2} E^T p_{\mathcal{E}} - k u_{\text{agc}}). \quad (30)$$

We define  $z = (\omega, p_{\mathcal{E}}, u_{\text{agc}})$  wherein  $\omega$  be the primal and  $(p_{\mathcal{E}}, u_{\text{agc}})$  be the dual variables. The variable vector  $z$  represents the dimensionless physical states: frequency, line powers, and AGC efforts. Matrix  $E$  describes the network topology,  $B$  the rescaled line susceptances,  $D$  the rescaled damping ratios, and  $k$  the secondary control gains. The PD form  $\dot{z} = (-\partial_{\omega} \ell, \partial_{p_{\mathcal{E}}} \ell, \partial_{u_{\text{agc}}} \ell)$  associated with the Lagrangian function (30) recovers frequency dynamics and the simplified AGC [14], [25].

For the perturbed system, we consider an extension of the model with additional turbine delays modeled as  $\hat{u}_{\text{agc}} = \frac{1}{T}(u_{\text{agc}} - \hat{u}_{\text{agc}})$ . Such delays play an important role in frequency transients and should be considered in any practical studies.

In this formulation, the frequency dynamics responds to the  $\hat{u}_{\text{agc}}$ , rather than the AGC effort  $u_{\text{agc}}$ . The resulting frequency equations becomes  $\dot{\omega} = -\partial_{\omega} \ell + k(\hat{u}_{\text{agc}} - u_{\text{agc}})$ .

We implement AGC on a SMIB (single-machine infinity-bus system) which consists of a generator and a purely inductive line. Further, we exert on the machine a sinusoidal exogenous torque which represents persistent load changes. Figure 1 compares the actual response to the estimated bound.

## V. CONCLUDING REMARKS

In this letter, we establish the strict contraction of the PD algorithm, applicable to a class of optimization problems that arise in network flows and distributed optimization. Strict contraction allows us to characterize the performance and robustness of the PD dynamics with respect to common deviations from the “true” PD dynamics. In particular, we consider the case of imperfect observers and also derive recursive bounds for hierarchical multi-scale optimization systems.

While in this letter we restricted ourselves to systems with equality constraints, future work will include more general extensions to inequality constraints. Furthermore, saddle-node dynamics also appear naturally in Brayton-Moser potentials [27], [28], which suggests an additional path for future research. Application-wise, we plan to explore the effect of inductive line delays [28], [29], on the performance of secondary controls [5] in microgrids.

In recent years, the multiple-time-scale optimization perspective has also taken on increased importance in natural systems, particularly in the context of systems biology [30]–[32]. In the brain, multiple interactions between different functional levels occur on a broad range of time-scales, involving weak or strong feedback interactions between genes, transcription factors, synapse formation, and global long-range connectivity [33]. In such systems, general Darwinism [21], [34] can play the role of an optimization criterion at every level [33], constrained by factors such as energy availability. Applying the tools developed in this letter to system modeling in such contexts is an additional subject for further research.

## REFERENCES

- [1] H. A. Simon, “The architecture of complexity,” in *Facets of Systems Science*. Boston, MA, USA: Springer, 1991, pp. 457–476.
- [2] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, “Layering as optimization decomposition: A mathematical theory of network architectures,” *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.
- [3] N. Li, C. Zhao, and L. Chen, “Connecting automatic generation control and economic dispatch from an optimization view,” *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 3, pp. 254–264, Sep. 2016.
- [4] C. Zhao, U. Topcu, N. Li, and S. Low, “Design and stability of load-side primary frequency control in power systems,” *IEEE Trans. Autom. Control*, vol. 59, no. 5, pp. 1177–1189, May 2014.
- [5] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, “Breaking the hierarchy: Distributed control and economic optimality in microgrids,” *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 3, pp. 241–253, Sep. 2016.
- [6] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-linear Programming*. Stanford, CA, USA: Stanford Univ. Press, 1958.
- [7] D. Feijer and F. Paganini, “Stability of primal–dual gradient dynamics and applications to network optimization,” *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
- [8] E. Mallada, C. Zhao, and S. Low, “Optimal load-side control for frequency regulation in smart grids,” *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6294–6309, Dec. 2017.
- [9] J. W. Simpson-Porco, “Input/output analysis of primal-dual gradient algorithms,” in *Proc. 54th Annu. Allerton Conf. IEEE Commun. Control Comput. (Allerton)*, 2016, pp. 219–224.
- [10] C. Zhao, E. Mallada, S. Low, and J. Bialek, “A unified framework for frequency control and congestion management,” in *Proc. IEEE Power Syst. Comput. Conf. (PSCC)*, 2016, pp. 1–7.
- [11] K. C. Kosaraju, S. Mohan, and R. Pasumarthy, “On the primal-dual dynamics of support vector machines,” *arXiv:1805.00699 [cs.SY]*, May 2018.
- [12] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards deep learning models resistant to adversarial attacks,” *arXiv:1706.06083 [stat.ML]*, Jun. 2017.
- [13] J. Wang and N. Elia, “A control perspective for centralized and distributed convex optimization,” in *Proc. 50th IEEE Conf. Decis. Control Eur. Control Conf.*, Orlando, FL, USA, Dec. 2011, pp. 3800–3805.
- [14] J. W. Simpson-Porco, B. K. Poolla, N. Monshizadeh, and F. Dörfler, “Quadratic performance of primal-dual methods with application to secondary frequency control of power systems,” in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Las Vegas, NV, USA, Dec. 2016, pp. 1840–1845.
- [15] A. Cherukuri, E. Mallada, S. Low, and J. Cortés, “The role of convexity on saddle-point dynamics: Lyapunov function and robustness,” *IEEE Trans. Autom. Control*, to be published.
- [16] W. Lohmiller and J.-J. E. Slotine, “On contraction analysis for non-linear systems,” *Automatica*, vol. 34, no. 6, pp. 683–696, 1998.
- [17] J.-J. E. Slotine, “Modular stability tools for distributed computation and control,” *Int. J. Adapt. Control Signal Process.*, vol. 17, no. 6, pp. 397–416, 2003.
- [18] D. Del Vecchio and J.-J. E. Slotine, “A contraction theory approach to singularly perturbed systems,” *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 752–757, Mar. 2013.
- [19] G. D. Bousquet and J.-J. E. Slotine, “A contraction based, singular perturbation approach to near-decomposability in complex systems,” *arXiv:1512.08464 [math.DS]*, Dec. 2015.
- [20] J.-J. E. Slotine *et al.*, *Applied Nonlinear Control*, vol. 60. Englewood Cliffs, NJ, USA: Prentice-Hall, 1991.
- [21] J.-P. Changeux, P. Courrège, and A. Danchin, “A theory of the epigenesis of neuronal networks by selective stabilization of synapses,” *Proc. Nat. Acad. Sci. USA*, vol. 70, no. 10, pp. 2974–2978, 1973.
- [22] W. Wang and J.-J. E. Slotine, “On partial contraction analysis for coupled nonlinear oscillators,” *Biol. Cybern.*, vol. 92, no. 1, pp. 38–53, 2005.
- [23] N. Tabareau, J.-J. Slotine, and Q.-C. Pham, “How synchronization protects from noise,” *PLoS Comput. Biol.*, vol. 6, no. 1, 2010, Art. no. e1000637.
- [24] G. Russo and J. J. E. Slotine, “Global convergence of quorum-sensing networks,” *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 82, no. 4, 2010, Art. no. 041919.
- [25] P. Kundur, *Power System Stability and Control* (EPRI Power System Engineering Series). New Haven, CT, USA: McGraw-Hill, 1994.
- [26] J. W. Simpson-Porco *et al.*, “Secondary frequency and voltage control of islanded microgrids via distributed averaging,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 11, pp. 7025–7038, Nov. 2015.
- [27] R. Ortega, D. Jeltsema, and J. M. A. Scherpen, “Power shaping: A new paradigm for stabilization of nonlinear RLC circuits,” *IEEE Trans. Autom. Control*, vol. 48, no. 10, pp. 1762–1767, Oct. 2003.
- [28] K. Cavanagh, J. A. Belk, and K. Turitsyn, “Transient stability guarantees for ad hoc DC microgrids,” *IEEE Control Syst. Lett.*, vol. 2, no. 1, pp. 139–144, Jan. 2018.
- [29] P. Vorobev, P.-H. Huang, M. A. Hosani, J. L. Kirtley, and K. Turitsyn, “High-fidelity model order reduction for microgrids stability assessment,” *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 874–887, Jan. 2018.
- [30] M. Kirschner and J. Gerhart, *The Plausibility of Life: Resolving Darwin’s Dilemma*. New Haven, CT, USA: Yale Univ. Press, 2005.
- [31] N. Kashtan and U. Alon, “Spontaneous evolution of modularity and network motifs,” *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 39, pp. 13773–13778, 2005.
- [32] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag, “Modular cell biology: Retroactivity and insulation,” *Mol. Syst. Biol.*, vol. 4, no. 1, p. 161, 2008.
- [33] J.-P. Changeux, “Climbing brain levels of organisation from genes to consciousness,” *Trends Cognit. Sci.*, vol. 21, no. 3, pp. 168–181, 2017.
- [34] G. M. Edelman, *Neural Darwinism: The Theory of Neuronal Group Selection*. New Haven, CT, USA: Basic Books, 1987.