# Exponential Stability of Partial Primal-Dual Gradient Dynamics with Nonsmooth Objective Functions

Zhaojian Wang<sup>a</sup>, Wei Wei<sup>a</sup>, Changhong Zhao<sup>b</sup>, Zetian Zheng<sup>a</sup>, Yunfan Zhang<sup>a</sup>, Feng Liu<sup>a,\*</sup>

<sup>a</sup>State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China <sup>b</sup>Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong

## Abstract

In this paper, we investigate the continuous time partial primal-dual gradient dynamics (P-PDGD) for solving convex optimization problems with the form  $\min_{x \in X, y \in \Omega} f(x) + h(y)$ , *s.t.* Ax + By = C, where f(x) is strongly convex and smooth, but h(y) is strongly convex and non-smooth. Affine equality and set constraints are included. We prove the exponential stability of P-PDGD, and bounds on decaying rates are provided. Moreover, it is also shown that the decaying rates can be regulated by setting the stepsize.

Keywords: Nonsmooth optimization, partial primal-dual gradient dynamics, Clark generalized gradient, exponential stability

### 1. Introduction

The primal-dual gradient dynamics (PDGD) (or saddle–point dynamics) were first introduced in Arrow et al. (1958); Kose (1956) and have been widely used in obtaining the primal-dual solutions of a convex (or concave) optimization problem. The partial primal-dual gradient dynamics (P-PDGD) is similar to PDGD, which is first proposed in Li et al. (2016) to solve specific types of optimization problems with separable decision variables, i.e., min f(x) + h(y), s.t. Ax + By = C. Typical applications of PDGD and P-PDGD include power systems Zhao et al. (2014); Li et al. (2016); Mallada et al. (2017); Wang et al. (2019a), wireless communication Chiang et al. (2007), distributed optimization Yi et al. (2015) and seeking the Nash Equilibrium in game Gharesifard and Cortés (2013).

Despite its wide applications, general theoretical studies of PDGD and P-PDGD are focused on its asymptotic stability analysis Feijer and Paganini. (2010); Cherukuri et al. (2016, 2018); Holding and Lestas (2014) and exponential stability analysis Cortés and Niederländer (2019); Niederländer et al. (2016); Dhingra et al. (2018); Qu and Li (2018); Tang et al. (2019); Chen and Li (2019); Liang et al. (2019); Bansode et al. (2019). In the existing literature, the global asymptotic stability of the PDGD under different settings is investigated. In Feijer and Paganini. (2010), the projection is proposed to address the inequality constraints and the PDGD is modeled as a hybrid automaton. This will result in the discontinuity of the righthand side of primal-dual dynamics and bring difficulties in the convergence proof. Then, Cherukuri et al. (2016) improves the convergence proof by using the invariance principle for discontinuous Caratheodory systems, which are further extended in

\*Corresponding author

Cherukuri et al. (2018) by using a weaker assumption to show global asymptotic stability for locally strong convex-concave Lagrangian. Extensions are given to subgradient methods in Holding and Lestas (2014), which constrain the dynamics to a convex domain.

Exponential stability is a desirable property of a dynamic system both theoretically and in practice. In many practical systems like the power grid, it is better to have stronger stability guarantees. In addition, a discrete-time iterative algorithm can be obtained from an exponentially stable dynamics through explicit Euler discretization, which achieves linear convergence for sufficiently small step sizes Stuart (1994); Stetter (1973). For PDGD to solve constrained convex optimization, its locally exponential stability can be obtained by investigating spectral bounds of saddle matrices Benzi et al. (2005). Regarding the global exponential stability, in Qu and Li (2018), the augmented PDGD is introduced to solve convex optimization with affine inequality constraints, where the exponential stability is proved. The method is further extended in Tang et al. (2019) for convex optimization problems with convex inequality constraints, and in Chen and Li (2019) for convex optimization problems with partially nonstrongly convex objective functions. In Liang et al. (2019), PDGD is used to solve the distributed optimization problem with nonstrongly convex objectives, where the metrically subregular condition is adopted to prove the exponential stability. In Bansode et al. (2019), a projected dynamics is proposed to solve a convex optimization problem with set and linear inequality constraints, where the exponential stability is proved on a Riemannian manifold. To solve nonsmooth convex optimization problems, Clark generalized gradients are adopted to replace the regular gradients. In Niederländer et al. (2016); Cortés and Niederländer (2019), saddle-point-like dynamics is proposed and proved to be globally exponentially stable with equality constraints. It is further improved in Dhingra et al. (2018) to consider affine inequality constraints.

<sup>&</sup>lt;sup>☆</sup>This work was supported by the National Natural Science Foundation of China (No. 51677100, U1966601, U1766206).

Email address: lfeng@mail.tsinghua.edu.cn (Feng Liu)

This work studies the exponential stability of P-PDGD to solve optimization problems with strongly convex and nonsmooth objective functions. Affine equality and set constraints are included. Compared with existing literature Cortés and Niederländer (2019); Niederländer et al. (2016); Dhingra et al. (2018); Qu and Li (2018); Tang et al. (2019); Chen and Li (2019); Liang et al. (2019), the main difference is that set constraints are considered. This is very important in practice. In many cases, set constraints are hard limits and even need to be satisfied during transient process. To this end, the projection onto the tangent cone of the set is adopted. This is also different from Bansode et al. (2019), where the projection onto the set itself is used. However, the method in Bansode et al. (2019) does not apply to optimization problems with nonsmooth objectives. As pointed out in Zeng et al. (2018), the projection of a subdifferential set cannot guarantee the existence of trajectories. This paper is also partially motivated by Li et al. (2020), which designs an exponentially convergent algorithm for the consensus problem. In this work, the optimization problem is more general and dual variables are explicitly included.

*Notations*: In this paper, use  $\mathbb{R}^n$  to denote the *n*-dimensional Euclidean space. For a column vector  $x \in \mathbb{R}^n$  (matrix  $A \in \mathbb{R}^{m \times n}$ ),  $x^{\mathrm{T}}(A^{\mathrm{T}})$  denotes its transpose. For vectors  $x, y \in \mathbb{R}^n$ ,  $x^{\mathrm{T}}y = \langle x, y \rangle$  denotes the inner product of x, y.  $||x|| = \sqrt{x^{\mathrm{T}}x}$  denotes the Euclidean norm of x. For any real symmetric matrices P and Q,  $P \geq Q$  and  $Q \leq P$  mean that P - Q is positive semidefinite.

### 2. Preliminaries

#### 2.1. Convex analysis

Let  $\Omega \subset \mathbb{R}^n$  be a convex set. Let  $f(x) : \Omega \to \mathbb{R}$  be a locally Lipschitz continuous function and denote its Clarke generalized gradient by  $\partial f(x)$  (Clarke, 1990, Page 27). For a  $\mu$ -strongly convex function  $f(x) : \mathbb{R}^n \to \mathbb{R}$ , we have  $(g_x - g_y)^T(x - y) \ge \mu ||x - y||^2$ ,  $\forall x, y \in \Omega$ , for all  $g_x \in \partial f(x)$  and  $g_y \in \partial f(y)$ .

For  $x \in \Omega$ , the tangent cone to  $\Omega$  at x is defined as

$$\mathcal{T}_{\Omega}(x) \triangleq \left\{ \lim_{k \to +\infty} \frac{x_k - x}{\tau_k} | x_k \in \Omega, x_k \to x, \tau_k > 0, \tau_k \to 0 \right\}$$

The normal cone to  $\Omega$  at *x* is defined by

$$\mathcal{N}_{\Omega}(x) = \{ v | \langle v, y - x \rangle \le 0, \forall y \in \Omega \}$$

By Brogliato et al. (2006), the tangent cone is the polar of the normal cone, which implies

$$T_{\Omega}(x) = \{ y \in \mathbb{R}^n | \langle s, y \rangle \le 0, \forall s \in N_{\Omega}(x) \}$$

Define the projection of x onto  $\Omega$  as

$$\mathcal{P}_{\Omega}(x) = \arg\min_{y \in \Omega} \|x - y\| \tag{1}$$

We have

$$x - \mathcal{P}_{\Omega}(x) \in \mathcal{N}_{\Omega}(x) \tag{2}$$

By Brogliato et al. (2006), the projection onto  $\mathcal{T}_{\Omega}(x)$  is computed by

$$P_{T_{\Omega}(x)}(y) = \lim_{\delta \to 0} \frac{P_{\Omega}(x + \delta y) - x}{\delta}$$
$$= y - \varpi z^{*}$$
(3)

where  $\varpi = \max \{0, \langle y, z^* \rangle\}$ , and  $z^* = \arg \max_{\|z\|=1, z \in \mathcal{N}_{\Omega}(x)} \langle y, z \rangle$ .

## 2.2. Differential inclusion

Following Aubin and Cellina (1984), a differential inclusion is given by

$$\dot{x}(t) \in \mathcal{F}(x(t)), x(0) = x_0, t \ge 0$$
 (4)

where  $\mathcal{F}$  is a set-valued map from points in  $\mathbb{R}^n$  to a nonempty, compact, convex subset of  $\mathbb{R}^n$ . For each  $x \in \mathbb{R}^n$ , system (4) specifies a set of possible evolutions instead of a single one. A solution of (4) defined on  $[0,\tau] \subset [0,\infty)$  is an absolutely continuous function  $x : [0,\tau] \to \mathbb{R}^n$  such that (4) holds for almost all  $t \in [0,\tau]$  for  $\tau > 0$ . The solution  $t \mapsto x(t)$  to (4) is a right maximal solution if it cannot be extended in time. Suppose that all the right maximal solutions to (4) exists on  $[0, +\infty)$ . If  $0_m \in \mathcal{F}(x^*)$ , then  $x^*$  is an equilibrium point of (4).

By (Aubin and Cellina, 1984, Proposition 2 in p. 266, and Theorem in p. 267), we have

**Lemma 1.** Let  $\Omega$  be a closed convex subset of  $\mathbb{R}^n$ , and  $\mathcal{F}$  be a map with non-empty compact value from  $\Omega$  to  $\mathbb{R}^n$ . Consider two differential inclusions given by

$$\dot{x}(t) \in \mathcal{F}(x(t)) - \mathcal{N}_{\Omega}(x(t)), x(0) = x_0$$
(5a)

$$\dot{x}(t) \in P_{\mathcal{T}_{\Omega}}[\mathcal{F}(x(t))], \quad x(0) = x_0 \tag{5b}$$

The trajectory x(t) is a solution of (5a) if and only if it is a solution of (5b). Moreover, if  $\mathcal{F}$  is also upper semi-continuous and bounded, there exists a solution to dynamics (5a).

Let  $V : \mathbb{R}^n \to \mathbb{R}$  be a locally Lipschitz continuous function and  $\partial V$  be the Clarke generalized gradient of V(x) at x. The setvalued Lie derivative  $\mathcal{L}_{\mathcal{F}}V$  of V with respect to (4) is defined as  $\mathcal{L}_{\mathcal{F}}V(x) \triangleq \{a \in \mathbb{R} : a = p^{\mathrm{T}}v, p \in \partial V(x), v \in \mathcal{F}(x)\}.$ 

#### 3. Problem Formulation and Algorithm Design

#### 3.1. Optimization problem

The optimization problem is

$$\min_{x \in X, y \in \Omega} \quad f(x) + h(y) \tag{6a}$$

s.t. 
$$Ax + By = C$$
 (6b)

where  $X \subset \mathbb{R}^n$ ,  $\Omega \subset \mathbb{R}^m$  are compact convex sets.  $A \in \mathbb{R}^{p \times n}$ ,  $B \in \mathbb{R}^{p \times m}$ ,  $C \in \mathbb{R}^p$  are constant matrices. We make following assumptions.

Assumption 1. The Slater's condition (Boyd and Vandenberghe, 2004, Chapter 5.2.3) of (6) holds.

Assumption 2. For some  $\alpha > 0$ , f(x) is  $\alpha$ -strongly convex and twice differentiable, i.e.,  $\nabla^2 f(x) \ge \alpha I$ .

Assumption 3. The function h(y) is Lipschitz continuous, and  $\beta$ -strongly convex on  $\Omega$  for some  $\beta > 0$ , that is,  $\langle y_1 - y_2, g_h(y_1) - g_h(y_2) \rangle \ge \beta ||y_1 - y_2||^2$ ,  $\forall y_1, y_2 \in \Omega$ , where  $g_h(y_1) \in \partial h(y_1)$  and  $g_h(y_2) \in \partial h(y_2)$ .

It should be noted that h(y) could be nonsmooth.

Assumption 4. The matrix A has full row rank and  $\kappa_1 I \leq AA^T$  for some  $\kappa_1 > 0$ .

**Remark 1.** Many practical problems have the same type as problem (6). For example, distributed frequency control problem in power systems Li et al. (2016); Wang et al. (2019b); Wang et al. (2019a), and distributed voltage control in distribution networks Liu et al. (2018); Wang et al. (2019b) can be generalized as (6). Thus, the result in this paper can be applied to many problems in practice.

#### 3.2. Algorithm Design

The Lagrangian of (6) is

$$L(x, y, \lambda) = f(x) + h(y) + \lambda^{\mathrm{T}}(Ax + By - C)$$
(7)  
$$_{x \in X, y \in \Omega}$$

where  $\lambda \in \mathbb{R}^p$  is the Lagrangian multiplier vector. The partial primal-dual gradient algorithm is

$$x = \arg\min_{x \in X} \left\{ f(x) + \lambda^{\mathrm{T}} A x \right\}$$
(8a)

$$\dot{y} \in P_{\mathcal{T}_{\Omega}(y)}\left(-\partial h\left(y\right) - B^{\mathrm{T}}\lambda\right)$$
(8b)

$$\dot{\lambda} = Ax + By - C \tag{8c}$$

Define functions

$$\varphi(\lambda) \triangleq \min_{x \in Y} \left\{ f(x) + \lambda^{\mathrm{T}} A x \right\}$$
(9)

$$\hat{L}(y,\lambda) \triangleq \varphi(\lambda) + h(y) + \lambda^{\mathrm{T}}(By - C)$$
(10)

Let  $H := \nabla^2 f(x)$ , and its maximal eigenvalue be  $\alpha_m$ . Its inverse  $H^{-1}$  is also positive definite, and the minimal eigenvalue is  $\frac{1}{\alpha_m}$ . Then, we have the following result demonstrating properties of  $\varphi(\lambda)$  and  $\hat{L}(y, \lambda)$ .

**Lemma 2.** Suppose Assumptions 2, 4 hold. The function  $\varphi(\lambda)$  is continuously differentiable and  $\frac{\kappa_1}{\alpha_m}$ -strongly concave. As a consequence,  $\hat{L}(y, \lambda)$  is strongly concave on  $\lambda$ .

*Proof.* Because f(x) is differentiable,  $\varphi(\lambda)$  is continuously differentiable by (Bertsekas., 2008, Proposition 6.1.1). The gradient of  $\varphi(\lambda)$  is  $\nabla \varphi(\lambda) = Ax$ . From (Bertsekas., 2008, Equation (6.9)), the Hessian of  $\varphi(\lambda)$  is

$$\nabla^2 \varphi(\lambda) = -AH^{-1}A^{\mathrm{T}} \tag{11}$$

For any  $z \in \mathbb{R}^p$ , we have

$$-z^{\mathrm{T}} \nabla^{2} \varphi(\lambda) z = z^{\mathrm{T}} A H^{-1} A^{\mathrm{T}} z$$

$$\geq \frac{1}{\alpha_{m}} z^{\mathrm{T}} A A^{\mathrm{T}} z$$

$$\geq \frac{\kappa_{1}}{\alpha_{m}} z^{\mathrm{T}} z \qquad (12)$$

where the second inequality is due to Assumption 4. This implies that  $-\varphi(\lambda)$  is  $\frac{\kappa_1}{\alpha_m}$ -strongly convex, i.e.,  $\varphi(\lambda)$  is  $\frac{\kappa_1}{\alpha_m}$ -strongly concave. This completes the proof.

From Lemma 2, taking any  $\lambda_1, \lambda_2 \in \mathbb{R}^p$ , we have

$$\langle \lambda_1 - \lambda_2, \nabla \varphi(\lambda_1) - \nabla \varphi(\lambda_2) \rangle \le -\frac{\kappa_1}{\alpha_m} \|\lambda_1 - \lambda_2\|^2$$
 (13)

Then, the algorithm (8) is rewritten as **P-PDGD:** 

$$x = \arg\min_{x \in Y} \left\{ f(x) + \lambda^{\mathrm{T}} A x \right\}$$
(14a)

$$\dot{y} \in P_{\mathcal{T}_{\Omega}(y)}\left(-\partial h(y) - B^{\mathrm{T}}\lambda\right)$$
 (14b)

$$\hat{A} = \nabla \varphi(\lambda) + By - C$$
 (14c)

In the rest of the paper, we will study the properties of the algorithm (14).

**Remark 2.** The algorithm (14) has some significant improvement from the initial version in Li et al. (2016). First, it considers the nonsmooth objective functions, where the Clark generalized gradient is utilized. Second, the set constraint is included, where the projection onto the tangent cone is adopted. Thus, the proposed algorithm is more general.

In (14b), the projection onto  $T_{\Omega}(y)$  is applied. The reason is that the projection from  $\partial h(y)$  onto  $\Omega$  may be a nonconvex differential inclusion. As a result, the existence of trajectories of (14) is not guaranteed, which sets a difficult barrier for convergence analysis.

#### 4. Optimality and Exponential Convergence

In this section, we first investigate the optimality of the equilibrium point of (14). Then, we prove that the trajectory of dynamics (14) converges to the equilibrium point exponentially.

#### 4.1. Optimality

Before proving the optimality, we introduce the existence of solutions to (14).

**Lemma 3.** Suppose Assumptions 2, 3, and 4 hold. Consider dynamics (14).

- 1. If  $y(0) \in \Omega$ , then  $y(t) \in \Omega$  for all  $t \ge 0$ .
- 2. If  $\lambda(0)$  is bounded, then  $\lambda(t)$  is bounded for all  $t \ge 0$ .
- 3. There exists a solution to (14).

Proof. For 1), define a function

$$E(y(t)) = \frac{1}{2} ||y(t) - P_{\Omega}(y(t))||^2$$
(15)

Its gradient is (Facchinei and Pang, 2003, Theorem 1.5.5)

$$\nabla E(\mathbf{y}(t)) = \mathbf{y}(t) - P_{\Omega}(\mathbf{y}(t)) \tag{16}$$

The time derivative of E(y(t)) is

$$\dot{E}(y(t)) = \langle \nabla E(y(t)), \dot{y}(t) \rangle = \langle y(t) - P_{\Omega}(y(t)), \dot{y}(t) \rangle$$
(17)

Because  $y(t) - P_{\Omega}(y(t)) \in \mathcal{N}_{\Omega}(y(t))$  and  $\dot{y}(t) \in \mathcal{T}_{\Omega}(y(t))$ , we have  $\dot{E}(y(t)) \leq 0$ . This implies that E(y(t)) is non-increasing. From  $E(y(t)) \geq 0$ ,  $\forall t \geq 0$  and E(y(0)) = 0, we have E(y(t)) = 0,  $\forall t \geq 0$ , i.e.,  $y(t) = P_{\Omega}(y(t))$ ,  $\forall t \geq 0$ .

For 2), define a function

$$\tilde{E}(\lambda(t)) = \frac{1}{2} \|\lambda(t)\|^2$$
(18)

The time derivative of  $\tilde{E}(\lambda(t))$  along (14) is

$$\dot{\tilde{E}}(\lambda) = \lambda^{\mathrm{T}}(\nabla\varphi(\lambda) + By - C)$$

$$= \lambda^{\mathrm{T}}\nabla\varphi(\lambda) + \lambda^{\mathrm{T}}(By - C)$$

$$\leq -\frac{\kappa_{1}}{\alpha_{m}}||\lambda||^{2} + \lambda^{\mathrm{T}}\nabla\varphi(0) + \lambda^{\mathrm{T}}(By - C)$$

$$\leq -\frac{\kappa_{1}}{\alpha_{m}}||\lambda||^{2} + a_{\lambda}||\lambda||$$

$$= -\frac{2\kappa_{1}}{\alpha_{m}}\tilde{E}(\lambda) + a_{\lambda}\sqrt{2\tilde{E}(\lambda)}$$
(19)

where  $a_{\lambda} = \max_{y \in \Omega} (\|\nabla \varphi(0)\| + \|By - C\|)$ , and the first inequality is due to  $\langle \lambda, \nabla \varphi(\lambda) \rangle \leq -\frac{\kappa_1}{\alpha_m} \|\lambda\|^2 + \lambda^T \nabla \varphi(0)$ . The second inequality is due to the boundedness of y(t). Then, we have  $\|\tilde{E}(\lambda(t))\| \leq \max\left\{\tilde{E}(\lambda(0)), \frac{\alpha_m^2 a_{\lambda}^2}{2\kappa_1^2}\right\}$ . Thus,  $\tilde{E}(\lambda(t)), t \geq 0$  is bounded, so is  $\lambda(t), t \geq 0$ .

Because h(y) is Lipschitz,  $\partial h(y)$  is nonempty, compact, convex, and upper semicontinuous (Cortes, 2008, Proposition 6). From 1) and 2), we have  $(y(t), \lambda(t))$  is bounded. Then, by Lemma 1, we can prove 3).

This completes the proof.

**Remark 3.** Lemma 3 shows that the trajectory of y(t) will stay in  $\Omega$  as long as  $y(0) \in \Omega$ . This is very important in practice besides paving the way for convergence proof. Many domain constraints are hard limits, which cannot be violated even in the transient process Trip et al. (2019). For example, the power generation limits of generators and capacity limits of inverters cannot be violated physically Wang et al. (2019b); Wang et al. (2019a). The voltage limits should not be violated, otherwise, it is dangerous for system operators Wang et al. (2019a). Thus, the results in the paper can be applied to many practical problems.

Let  $(x^*, y^*, \lambda^*)$  be an equilibrium of (14). Then

$$x^* = \arg\min_{x \in \mathbf{V}} \left\{ f(x) + (\lambda^*)^{\mathrm{T}} A x \right\}$$
(20a)

$$0 \in P_{\mathcal{T}_{\Omega}(y^*)}\left(-\partial h\left(y^*\right) - B^{\mathrm{T}}\lambda^*\right)$$
(20b)

$$0 = Ax^* + By^* - C$$
 (20c)

**Theorem 4.** Suppose Assumptions 1, 2, 3 hold. The point  $(x^*, y^*, \lambda^*)$  satisfies (20), if and only if it is the primal-dual optimal solution to (6) and its dual problem.

*Proof.* According to the Karush-Kuhn-Tucker (KKT) optimal conditions (Ruszczyński and Ruszczynski, 2006, Theorem 3.25), the primal-dual optimal solution should satisfy

$$0 \in \nabla f(x^*) + A^{\mathrm{T}}\lambda^* + \mathcal{N}_X(x^*)$$
(21a)

$$0 \in \partial h(y^*) + B^{\mathrm{T}}\lambda^* + \mathcal{N}_{\Omega}(y^*)$$
(21b)

$$0 = Ax^* + By^* - C$$
 (21c)

Compare (21) with (20), and we know (21a) is equivalent to (20a). By Lemma 1, we have (21b) is equivalent to (20b). Thus, (21) is equivalent to (20). Because the optimization problem (6) is convex and with strongly convex objective functions,  $(x^*, y^*, \lambda^*)$  is the primal-dual optimal solution to (6) and its dual problem. This completes the proof.

#### 4.2. Exponential convergence

In this subsection, we analyze the convergence rate for algorithm dynamics (14).

**Theorem 5.** Suppose Assumptions 1, 2, 3 hold. The solution algorithm dynamics (14) converges to its equilibrium point  $(x^*, y^*, \lambda^*)$  exponentially.

Proof. Define the Lyapunov function candidate

$$V(y,\lambda) = \frac{1}{2} ||y - y^*||^2 + \frac{1}{2} ||\lambda - \lambda^*||^2$$
(22)

The time derivative of  $V(y, \lambda)$  is

$$\mathcal{L}_{\mathcal{F}}V(y,\lambda) = \{a \in \mathbb{R} : \\ a = \nabla_{y}V(y,\lambda)^{\mathrm{T}}P_{\mathcal{T}_{\Omega}(y(t))}\left(-\partial h\left(y(t)\right) - B^{\mathrm{T}}\lambda\right) \\ + \nabla_{\lambda}V(y,\lambda)^{\mathrm{T}}(\nabla\varphi(\lambda) + By - C)\}$$
(23)

Suppose  $a \in \mathcal{L}_{\mathcal{F}} V(y, \lambda)$ . There is  $\eta(y) \in \partial h(y)$  such that

$$a = (y - y^*)^{\mathrm{T}} P_{\mathcal{T}_{\Omega}(y(t))} \left( -\eta(y) - B^{\mathrm{T}} \lambda \right) + (\lambda - \lambda^*)^{\mathrm{T}} (\nabla \varphi(\lambda) + By - C)$$
(24)

From (5a), we have

$$-\eta(y) - B^{\mathrm{T}}\lambda - \dot{y} \in \mathcal{N}_{\Omega}(y) \tag{25}$$

From the definition of  $N_{\Omega}(y(t))$  and the fact that  $y^* \in \Omega$ , we have

$$\left\langle y^* - y, -\eta(y) - B^{\mathrm{T}}\lambda - \dot{y} \right\rangle \le 0$$

$$\left\langle y - y^*, \dot{y} \right\rangle \le \left\langle y - y^*, -\eta(y) - B^{\mathrm{T}}\lambda \right\rangle$$
(26)

Then

$$a \le (y - y^*)^T \left(-\eta(y) - B^T \lambda\right) + (\lambda - \lambda^*)^T \left(\nabla \varphi(\lambda) + By - C\right)$$
(27)

From (21), the definition of  $N_{\Omega}(x)$ , and the fact that  $y(t) \in \Omega, \forall t$ , at the equilibrium, we have

$$\left\langle -\eta\left(y^{*}\right) - B^{\mathrm{T}}\lambda^{*}, y - y^{*}\right\rangle \le 0 \tag{28a}$$

$$0 = \nabla \varphi(\lambda^*) + By^* - C \tag{28b}$$

Combine (28) with (27), and we have

$$a \leq (y - y^{*})^{\mathrm{T}} \left( -(\eta(y) - \eta(y^{*})) - B^{\mathrm{T}}(\lambda - \lambda^{*}) \right) + (\lambda - \lambda^{*})^{T} (\nabla \varphi(\lambda) - \nabla \varphi(\lambda^{*}) + By - By^{*}) \leq -(y - y^{*})^{\mathrm{T}} (\eta(y) - \eta(y^{*})) + (\lambda - \lambda^{*})^{T} (\nabla \varphi(\lambda) - \nabla \varphi(\lambda^{*})) \leq -\beta ||y - y^{*}||^{2} - \frac{\kappa_{1}}{\alpha_{m}} ||\lambda - \lambda^{*}||^{2} \leq -\gamma V$$
(29)

where  $\gamma = \min \{2\beta, 2\frac{\kappa_1}{\alpha_m}\}$ . As a result,  $V(t) \leq V(0)e^{-\gamma t}$ , and we have  $||(y(t) - y^*, \lambda(t) - \lambda^*)|| \leq \sqrt{2V(0)}e^{-\frac{\gamma}{2}t}$  converges to  $(y^*, \lambda^*)$  exponentially, and convergence rate is no less than  $\frac{\gamma}{2}$ . This completes the proof.

**Remark 4** (Decaying rate). *Give some*  $\tau > 0$ , *the P-PDGD can be written as* 

$$x = \arg\min_{x \in X} \left\{ f(x) + \lambda^{\mathrm{T}} A x \right\}$$
(30a)

$$\dot{y} \in \tau P_{\mathcal{T}_{\Omega}(y)}\left(-\partial h(y) - B^{\mathrm{T}}\lambda\right)$$
 (30b)

$$\dot{\lambda} = \tau (\nabla \varphi(\lambda) + By - C)$$
 (30c)

Then, the Lyapunov function is

$$V_2(y,\lambda) = \frac{1}{2\tau} ||y - y^*||^2 + \frac{1}{2\tau} ||\lambda - \lambda^*||^2$$
(31)

Follow the similar analysis in Theorem 5, and we have

$$a \leq -\frac{1}{2}\gamma \left\| (y(t) - y^*, \lambda(t) - \lambda^*) \right\|^2$$
  
$$\leq -\gamma\tau V_2(y, \lambda)$$
(32)

As a result,  $V_2(t) \leq V_2(0)e^{-\gamma \tau t}$ , and we have  $||(y(t) - y^*, \lambda(t) - \lambda^*)|| \leq \sqrt{2\tau V_2(0)}e^{-\frac{\gamma \tau}{2}t}$ . Thus, the decaying rate bound can be regulated by  $\tau$ . If  $\tau = 1$ , it will be same as that in Theorem 5.

#### 5. Illustrative Examples

We consider a convex problem that models the optimal voltage control in distribution networks. The problem is formulated as

$$\min_{U,q \in \mathbb{R}^n} \quad f = \frac{a}{2} \|U - \mathbf{1}\|^2 + \sum_{j=1}^n h_j(q_j)$$
(33a)

s.t. 
$$BU = q + C$$
 (33b)

$$q \le q \le \overline{q} \tag{33c}$$

where U is the voltage, q is the reactive power. BU = q + Cis the linear model derived from the DistFlow equations Baran and Wu (1989).  $q, \bar{q}$  are lower and upper bound of q. a is a real positive constant. The first part of objective function,  $\frac{a}{2}||U - 1||^2$ , is the voltage difference, where 1 is the nominal voltage. The second part,  $h_j(q_j)$ , is the regulation cost of reactive power,



Figure 1: Dynamics of reactive power q at each bus

which is strongly convex and nonsmooth, defined by

$$h_j(q_j) = \begin{cases} q_j^2 - 0.02, & q_j \le -0.2\\ \frac{1}{2}q_j^2, & -0.2 < q_j \le 0.2\\ q_j^2 - 0.02, & 0.2 < q_j \end{cases}$$
(34)

An 8-bus feeder is utilized as the test system, the detailed description is given in Wang et al. (2019b). The bus 0 is the substation with voltage as 1 and has no generator. Other have reactive power regulation capability with  $\overline{q} = -q = [80, 80, 88, 80, 104, 80, 96]$ kVar. We set a = 8,  $C = (1.011, -0.009, -0.1, 0.14, -0.26, -0.019, -0.06)^{T}$ . The minimal eigenvalue of *B* is 0.1165. Moreover, we have  $\beta = 1, \kappa_1 = 0.1165^2 = 0.01357, \alpha_m = a = 8$ . Therefore  $\gamma = \min\{2\beta, 2\kappa_1/\alpha_m\} = \min\{2, 2 \times 0.01357/8\} = 0.00339$ .

Define  $\Omega \triangleq \{q : \underline{q} \le q \le \overline{q}\}$ , and then the P-PDGD for (33) is

$$U = \arg\min_{U} \left\{ \frac{a}{2} \|U - \mathbf{1}\|^2 + \lambda^{\mathrm{T}} B U \right\} = -\frac{1}{a} B^{\mathrm{T}} \lambda + \mathbf{1}$$
(35a)

$$\dot{q} \in P_{\mathcal{T}_{\Omega}(q)}\left(-\partial h\left(q\right) + \lambda\right)$$
 (35b)

$$\dot{\lambda} = -\frac{1}{a}BB^{\mathrm{T}}\lambda - q + B \times \mathbf{1} - C \tag{35c}$$

The method in (3) is utilized to compute the projection onto  $\mathcal{T}_{\Omega}(q)$ . The simulations are implemented on Matlab R2013b, and the function "ode23tb" is adopted to solve (35). The equilibrium point of (35) is

$$q^* = (0.031, 0.031, 0.042, 0.075, 0.104, 0.035, 0.044)^{\mathrm{T}},$$

$$U^* = (0.992, 1.003, 1.004, 1.031, 0.959, 1.004, 0.997)^{\mathrm{T}}.$$

The dynamic performance of (35) are given in Fig.1 and Fig.2. In Fig.1, the dynamics of reactive power are illustrated, which shows that  $q_i(t)$ ,  $\forall i$  is within its lower and upper limits. This validates Lemma 3.

In Fig.2, the convergence of  $||(q(t) - q^*, \lambda(t) - \lambda^*)||$ is illustrated. The dotted line is the trajectory of  $\sqrt{||q(0) - q^*||^2 + ||\lambda(0) - \lambda^*||^2}e^{-\gamma t/2}$ , which is the upper bound of exponential convergence obtained from Theorem 5. The blue line is the trajectory of  $(q(t), \lambda(t))$ . It is shown that  $(q(t), \lambda(t))$  converges to  $(q^*, \lambda^*)$  rapidly. The convergence speed is much faster than computed upper bound.

We further investigate the decaying rates under different  $\tau$ , which is illustrated in Fig.3. With the increasing of  $\tau$ , the de-



Figure 2: Illustration of the relative distances to  $(q^*, \lambda^*)$ 



Figure 3: Illustration of the decaying rate with different  $\tau$ . The dotted lines are trajectories of  $\sqrt{2\tau V_2(0)}e^{-\frac{\tau}{2}t}$  with different  $\tau$ . The solid lines are the corresponding trajectories of  $||(y(t) - y^*, \lambda(t) - \lambda^*)||$ .

caying rates also increase, and the exponential convergence always holds. The result is consistent with the analysis in Remark 4.

## 6. Conclusion

This paper investigates the P-PDGD for solving convex optimization with strongly convex and non-smooth objectives. Affine equality and set constraints are considered. We prove the exponential stability of P-PDGD, where bounds on decaying rates are also provided. It is also validated that the algorithm is almost initialization free as long as the initial point satisfies the set constraints.

Our results are promising in many practical problems, such as the frequency and voltage control in power systems, which can provide a stronger stability guarantee. However, there are still some limitations on the problem form. In the future, we will investigate exponentially convergent algorithms for more general optimization problems.

#### References

- Arrow, K.J., Hurwicz, L., Uzawa, H., 1958. Studies in linear and non-linear programming. Cambridge Univ. Press.
- Aubin, J.P., Cellina, A., 1984. Differential inclusions. Springer-Verlag.
- Bansode, P., Chinde, V., Wagh, S., Pasumarthy, R., Singh, N., 2019. On the exponential stability of projected primal-dual dynamics on a riemannian manifold. arXiv preprint arXiv:1905.04521.
- Baran, M., Wu, F.F., 1989. Optimal sizing of capacitors placed on a radial distribution system. IEEE Trans. Power Delivery 4, 735–743.

- Benzi, M., Golub, G.H., Liesen, J., 2005. Numerical solution of saddle point problems. Acta numerica 14, 1–137.
- Bertsekas., D.P., 2008. Nonlinear programming. 2 ed., Athena scientific.
- Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge university press.
- Brogliato, B., Daniilidis, A., Lemarechal, C., Acary, V., 2006. On the equivalence between complementarity systems, projected systems and differential inclusions. Systems & Control Letters 55, 45–51.
- Chen, X., Li, N., 2019. Exponential stability of primal-dual gradient dynamics with non-strong convexity. arXiv preprint arXiv:1905.00298.
- Cherukuri, A., Mallada, E., Corts, J., 2016. Asymptotic convergence of constrained primaldual dynamics. Syst. Control Lett. 87, 10 – 15.
- Cherukuri, A., Mallada, E., Low, S., Cortes, J., 2018. The role of convexity in saddle-point dynamics: Lyapunov function and robustness. IEEE Trans. Autom. Control 63, 2449–2464.
- Chiang, M., Low, S.H., Calderbank, A.R., Doyle, J.C., 2007. Layering as optimization decomposition: A mathematical theory of network architectures. Proc. IEEE 95, 255–312.
- Clarke, F.H., 1990. Optimization and nonsmooth analysis. volume 5. Siam.
- Cortes, J., 2008. Discontinuous dynamical systems. IEEE Control Syst. Mag. 28, 36–73.
- Cortés, J., Niederländer, S.K., 2019. Distributed coordination for nonsmooth convex optimization via saddle-point dynamics. Journal of Nonlinear Science 29, 1247–1272.
- Dhingra, N.K., Khong, S.Z., Jovanović, M.R., 2018. The proximal augmented Lagrangian method for nonsmooth composite optimization. IEEE Trans. Autom. Control 64, 2861–2868.
- Facchinei, F., Pang, J.S., 2003. Finite-dimensional variational inequalities and complementarity problems. Springer-Verlag, New York.
- Feijer, D., Paganini., F., 2010. Stability of primal-dual gradient dynamics and applications to network optimization. Automatica 46, 1974–1981.
- Gharesifard, B., Cortés, J., 2013. Distributed convergence to Nash equilibria in two-network zero-sum games. Automatica 49, 1683–1692.
- Holding, T., Lestas, I., 2014. On the convergence to saddle points of concaveconvex functions, the gradient method and emergence of oscillations, in: 53rd IEEE Conference on Decision and Control, IEEE. pp. 1143–1148.
- Kose, T., 1956. Solutions of saddle value problems by differential equations. Econometrica, Journal of the Econometric Society, 59–70.
- Li, N., Zhao, C., Chen, L., 2016. Connecting automatic generation control and economic dispatch from an optimization view. IEEE Trans. Control Network Syst. 3, 254–264.
- Li, W., Zeng, X., Liang, S., Hong, Y., 2020. Exponentially convergent algorithm design for constrained distributed optimization via non-smooth approach. arXiv preprint arXiv:2001.00509.
- Liang, S., Yin, G., et al., 2019. Exponential convergence of distributed primal– dual convex optimization algorithm without strong convexity. Automatica 105, 298–306.
- Liu, H.J., Shi, W., Zhu, H., 2018. Hybrid voltage control in distribution networks under limited communication rates. IEEE Transactions on Smart Grid 10, 2416–2427.
- Mallada, E., Zhao, C., Low, S., 2017. Optimal load-side control for frequency regulation in smart grids. IEEE Trans. Autom. Control 62, 6294–6309.
- Niederländer, S.K., Allgöwer, F., Cortés, J., 2016. Exponentially fast distributed coordination for nonsmooth convex optimization, in: 2016 IEEE 55th Conference on Decision and Control (CDC), IEEE. pp. 1036–1041.
- Qu, G., Li, N., 2018. On the exponential stability of primal-dual gradient dynamics. IEEE Control Systems Letters 3, 43–48.
- Ruszczyński, A.P., Ruszczynski, A., 2006. Nonlinear optimization. volume 13. Princeton university press.
- Stetter, H.J., 1973. Analysis of discretization methods for ordinary differential equations. volume 23. Springer.
- Stuart, A.M., 1994. Numerical analysis of dynamical systems. Acta numerica 3, 467–572.
- Tang, Y., Qu, G., Li, N., 2019. Semi-global exponential stability of primaldual gradient dynamics for constrained convex optimization. arXiv preprint arXiv:1903.09580.
- Trip, S., Scholten, T., De Persis, C., 2019. Optimal regulation of flow networks with transient constraints. Automatica 104, 141–153.
- Wang, Z., Liu, F., Chen, Y., Low, S.H., Mei, S., 2019a. Unified distributed control of stand-alone dc microgrids. IEEE Trans. Smart Grid 10, 1013– 1024.

- Wang, Z., Liu, F., Low, S.H., Zhao, C., Mei, S., 2019b. Distributed frequency control with operational constraints, part I: Per-node power balance. IEEE Trans. Smart Grid 10, 40–52.
- Wang, Z., Liu, F., Low, S.H., Zhao, C., Mei, S., 2019a. Distributed frequency control with operational constraints, part II: Network power balance. IEEE Trans. Smart Grid 10, 53–64.
- Wang, Z., Liu, F., Su, Y., Qin, B., 2019b. Asynchronous distributed voltage control in active distribution networks. arXiv preprint arXiv:1908.06548.
- Yi, P., Hong, Y., Liu, F., 2015. Distributed gradient algorithm for constrained optimization with application to load sharing in power systems. Syst. Control Lett. 83, 45–52.
- Zeng, X., Yi, P., Hong, Y., Xie, L., 2018. Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems. SIAM J. Control Optim. 56, 3973–3993.
- Zhao, C., Topcu, U., Li, N., H.Low., S., 2014. Design and stability of load-side primary frequency control in power systems. IEEE Trans. Autom. Control 59, 1177–1189.