Letter

Exponential Convergence of Primal-Dual Dynamical System for Linear Constrained Optimization

Luyao Guo, Xinli Shi, *Member, IEEE*, and Jinde Cao, *Fellow, IEEE*

Dear editor,

Primal-dual dynamics (PDD) and its variants are prominent first-order continuous-time algorithms to determine the primal and dual solutions of a constrained optimization problem (COP). Due to the simple structure, they have received widespread attention in various fields, such as distributed optimization [1], power systems [2], and wireless communication [3]. In view of their wide applications, there are numerous theoretic studies on the convergence properties of PDD and its variants, including the exponential stability analysis [4]–[9].

In [4], an asymptotically convergent PDD is proposed to solve the optimization problems with equality and inequality constraints. Especially for equality COPs, the exponential convergence of the proposed PDD is established under the assumption that the equality constraint matrix has a full row rank. For COPs with linear equality and inequality, PDDs with exponential convergence are respectively designed under the full row rank assumption in [5]. Besides, an extended augmented PDD is provided to solve linear equality and nonlinear inequality COPs, and the semi-global exponential stability is established in [6]. To solve a special class of COPs with separable cost function and equality constraint, [7] provides a Lagrangian-based PDD, and [8] gives a partial PDD. Moreover, with full row rank assumption, their exponential convergence are proved. In [9], a fixed-time convergent PDD has been proposed for equality COPs.

In this letter, we investigate a modified primal-dual dynamical system (PDDS) for COPs with linear equality and inequality constraints, and establish the exponential convergence of corresponding PDDS under weaker conditions. The obtained theoretical results without requiring a rank condition on equality constraint matrix can be used to improve the convergence results of [8] and [9].

Problem statement:

Consider the following COP.

$$f_{opt}=\min_{x\in\mathbb{R}^n} f(x), \qquad \text{s.t. } A_1x=b_1,\ A_2x\le b_2 \tag{1}$$

where $f:\ \mathbb{R}^n\to\mathbb{R}$ is a convex function, $A_1\in\mathbb{R}^{m_1\times n}$, $A_2\in\mathbb{R}^{m_2\times n}$, $b_1\in\mathbb{R}^{m_1}$ and $b_2\in\mathbb{R}^{m_2}$. Throughout the letter, we give the following assumption.

Assumption 1: The cost function $f(x)$ is μ_f-strongly convex with an l_f-Lipschitz continuous gradient. The optimal set X^* is nonempty and closed, and the optimal value f_{opt} is finite.

For problem (1), the Lagrangian is given as follows:

$$L(x,\lambda)=f(x)+\lambda^T(Hx-h) \tag{2}$$

Citation: L. Y. Guo, X. L. Shi, and J. D. Cao, "Exponential convergence of primal-dual dynamical system for linear constrained optimization," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 4, pp. 745–748, Apr. 2022.

L. Y. Guo is with the School of Mathematics, Southeast University, Nanjing 210096, China (e-mail: guo_luyaoo@163.com).

X. L. Shi is with School of Cyber Science & Engineering, Southeast University, Nanjing 210096, China (e-mail: xinli_shi@seu.edu.cn).

J. D. Cao is with School of Mathematics, Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 210096, with Purple Mountain Laboratories, Nanjing 211111, China, and also with Yonsei Frontier Laboratory, Yonsei University, Seoul 03722, South Korea (e-mail: jdcao@seu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2022.105485

where $\lambda=(\phi,\varphi)\in\mathbb{D}:=\mathbb{R}^{m_1}\times\mathbb{R}_+^{m_2}$ is the dual variable, $H=[A_1^T, A_2^T]^T$, and $h=[b_1^T,b_2^T]^T$. Assumption 1 implies that strong duality holds: $f_{opt}=\max_{\lambda\in\mathbb{D}} d(\lambda)$, where $d(\lambda)=\min_{x\in\mathbb{R}^n} L(x,\lambda)$ is the dual function of (1), and the optimal solution set Λ^* of the dual problem is nonempty. $(x^*,\lambda^*)\in X^*\times\Lambda^*$ is the pair of primal-dual optimal solution iff x^* is feasible and $x^*=\arg\min_x L(x,\lambda^*)$, which means that one can recover a primal optimal point x^* from a dual optimal point λ^* by solving the equation $\nabla f(x^*)+H^T\lambda^*=0$.

By Assumption 1 and the definition of conjugate function, $d(\lambda)$ is l_d-smooth, and $d(\lambda)=-f^*(-H^T\lambda)-h^T\lambda$, where f^* is the convex conjugate of f. However, $-d(\lambda)$ may not be strongly convex, because only when H has full row rank, $f^*(-H^T\lambda)$ is strongly convex. For example, if $f^*(y)=(1/2)\|y\|^2$ and H is row rank deficient, $f^*(-H^T\lambda)=(1/2)\|H^T\lambda\|^2$ is not strongly convex. Therefore, without strong convexity of the dual function $-d(\lambda)$, the establishment of the exponential convergence of PDDS for the problem (1) is naturally difficult. To deal with this difficulty, [4]–[9] assume that H has full row rank, and [10] limits the initial condition of the dual variable $\lambda(0)$ to the column space of H. Different from these studies, this letter neither assumes that H has full row rank nor restricts $\lambda(0)$ to the column space of H.

Main results: This section considers the exponential convergence of PDDS to solve the COP (1) and then consider the case with equality constraints alone.

COPs with equality and inequality constraints:

Consider the problem $\min\{f(x): A_1x=b_1, A_2x\le b_2\}$. Let $\mathcal{S}=\{x\in\mathbb{R}^n: A_1x=b_1, A_2x\le b_2\}$ denote the constraint set. Throughout this subsection, suppose that the set $\mathcal{S}\subset\mathbb{R}^n$ satisfies strong Slater assumptions, i.e.,

Assumption 2: A_1 has full row rank, and there exists a feasible point $\tilde{x}$ such that $A_1\tilde{x}=b_1$ and $A_2\tilde{x}<b_2$.

Under Assumptions 1 and 2, Λ^* is nonempty, convex, and bounded, which is particularly significant in the analysis of the convergence properties of the following projected dual gradient dynamics.

$$\begin{cases} x(t)=\arg\min\limits_{s\in\mathbb{R}^n} L(s,\lambda(t))\\ \dot{\lambda}(t)=-\lambda(t)+\mathcal{P}_{\mathbb{D}}\big(\lambda(t)+\upsilon\cdot\nabla d(\lambda(t))\big),\ \lambda(0)\in\mathbb{D}\end{cases} \tag{3}$$

where $0<\upsilon\le 1/l_d$, and $\mathcal{P}_{\mathbb{D}}(\lambda)$ denotes the projection of λ onto $\mathbb{D}$. By [11, Theorem 5.2], for any initial condition $\lambda(0)\in\mathbb{D}$, the solution of (3) exists globally. Moreover, any solution $\lambda(t)$ of (3) with $\lambda(0)\in\mathbb{D}$ is bounded, and satisfies that $\lambda(t)\in\mathbb{D}$ for all $t\ge 0$. Then, we give the following theorem to show the convergence properties of algorithms of (3), and provide the exponential convergence of $f_{opt}-d(\lambda(t))$, $\|\mathcal{P}_{\mathbb{D}}(Hx(t)-h)\|$, and $|f(x(t))-f_{opt}|$.

Theorem 1: Supposing that Assumptions 1 and 2 holds, the trajectory $\lambda(t)$ of (3) converges exponentially to Λ^* with any initial condition $\lambda(0)\in\mathbb{D}$. Furthermore, there exist $\alpha>0$, $\beta>0$ and $\gamma>0$ such that the following convergence properties hold.

$$f_{opt}-d(\lambda(t))\le\alpha\cdot e^{-\beta t}, \quad \forall t\ge 0 \tag{4}$$

$$\|\mathcal{P}_{\mathbb{D}}(Hx(t)-h)\|\le\sqrt{\frac{2\alpha}{\upsilon}}\cdot e^{-\frac{\beta}{2}t}, \quad \forall t\ge 0 \tag{5}$$

$$\|x(t)-x^*\|\le\sqrt{\frac{2\alpha}{\mu_f}}\cdot e^{-\frac{\beta}{2}t}, \quad t\ge 0 \tag{6}$$

$$|f(x(t))-f_{opt}|\le\gamma e^{-\frac{\beta}{2}t}, \quad t\ge 0. \tag{7}$$

Proof: See Appendix A. ■

Noting that if $-d(\lambda)$ is strongly convex, or if $-\nabla d(\lambda)$ is strongly pseudomonotone and Λ^* is a singleton, the projected dynamics (3) is exponentially convergent [11], [12]. In contrast, for the PDDS (3), we eliminate these assumptions and thus extend these results.

COPs with equality constraints:

In this subsection, without strong Slater assumptions, we consider the special case of (1), where $H = A_1^T$ and $h = b_1^T$. Here, the matrix A_1 is not required to have full row rank, and the PDDS (3) degenerates into the following form.

$$x(t) = \arg\min_s L(s,\lambda(t)) \tag{8a}$$

$$\dot{\lambda}(t) = \nabla d(\lambda(t)). \tag{8b}$$

The dual function of this problem is $d(\lambda) = -f^*(-A_1^T\lambda) - (x^*)^T A_1^T \lambda$, where $x^* \in X^*$. If A_1 is row rank deficient, the negative dual function $-d(\lambda)$ is not strongly convex. Moreover, Λ^* would be unbounded: for any $\lambda^* \in \Lambda^*$, the whole affine manifold $\lambda^* + \mathrm{Null}(A_1^T)$ would be in Λ^*, where $\mathrm{Null}(A_1^T) = \{\lambda : A_1^T\lambda = 0\}$. Therefore, the Theorem 1 is no longer applicable.

Next, we analyze the convergence of (8) under weaker conditions. By [13, Theorem 9], $-d(\lambda)$ has quadratic gradient growth, i.e.,

$$(\lambda - [\lambda]^*)^T \nabla(-d)(\lambda) \geq \frac{1}{(2l_f\omega)\|\lambda - [\lambda]^*\|^2} \tag{9}$$

where $[\lambda]^*$ is the projection of λ onto Λ^* and ω is a constant only related to A_1. Moreover, it can be shown that there exists a unique $s^* \in \mathbb{R}^n$ such that $A_1^T\lambda^* = s^*$ for any $\lambda^* \in \Lambda^*$, and the set Λ^* can be characterized as the following polyhedron: $\Lambda^* = \{\lambda : A_1^T \lambda = s^*\}$. Through the above analysis, we provide the following theorem to show that the PDDS (8) is globally exponentially convergent without strong Slater assumptions.

Theorem 2: Supposing that Assumption 1 holds, the trajectory $\lambda(t)$ of dynamics (8) converges exponentially to Λ^* for any $\lambda(0) \in \mathbb{R}^m$. Furthermore, there exist $\bar{\alpha}, \bar{\beta}, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma} > 0$ such that the following convergence properties hold.

$$f_{opt} - d(\lambda(t)) \leq \tilde{\alpha} \cdot e^{-\tilde{\beta}t}, \quad t \geq 0 \tag{10}$$

$$\|A_1 x(t) - b_1\| \leq l_d \bar{\alpha} \cdot e^{-\frac{\bar{\beta}}{2}t}, \quad t \geq 0 \tag{11}$$

$$\|x(t) - x^*\| \leq \sqrt{\frac{2\tilde{\alpha}}{\mu_f}} e^{-\frac{\tilde{\beta}}{2}t}, \quad t \geq 0 \tag{12}$$

$$|f(x(t)) - f_{opt}| \leq \tilde{\gamma} e^{-\frac{\tilde{\beta}}{2}t}, \quad t \geq 0. \tag{13}$$

Proof: See Appendix B. ■

From a practical perspective, the assumption of full row rank can not be sufficed easily, such as distributed optimization. For distributed optimization problems over an undirected and connected network, several primal-dual approaches have been provided in [14]–[17]. In these literature, by introducing the oriented incidence matrix $\mathcal{B}$ or the Laplacian matrix $\mathcal{L}$ of the considered network, the consensus constraints can be equivalently formulated as $\mathcal{B}^T x = 0$ or $\mathcal{L}x = 0$, which not only simplifies the consensus equality constraint, but also makes it can be handled in a distributed manner. However, both $\mathcal{B}$ and $\mathcal{L}$ are row rank deficient. Therefore, Theorem 2 is significant in theory and practice.

In the comparison, several PDDs provided in [4], [5], [7] for equality COPs employ the technique that the smooth part of $L(x,\lambda(t))$ is replaced by the quadratic approximation centered at $x(t-\tau)$, $\tau > 0$, i.e., $L(x,\lambda(t)) \approx (\nabla_x L(x(t-\tau),\lambda(t)))^T(x - x(t-\tau)) + (1/2\tau)\|x - x(t-\tau)\|^2$. And they have been proven exponentially convergent under full row rank assumption. In Theorem 2, we not only establish the exponential convergence of the PDDS (8) under weaker conditions, but also fully discuss its convergence properties.

Then, applying the analysis techniques in Theorem 2 to some existing primal-dual dynamics systems [8], [9], we establish their exponential and fixed-time convergence respectively under weaker conditions and improve their convergence results.

In [8], the following separable problem is considered.

$$\min_{x_1 \in \mathcal{X}, x_2 \in \Omega} f_1(x_1) + f_2(x_2), \quad \text{s.t.} \ A_1x_1 + A_2x_2 = C \tag{14}$$

where $x_1 \in \mathcal{X} \subset \mathbb{R}^{N_1}$, $x_2 \in \Omega \subset \mathbb{R}^{N_2}$, is closed and convex, $A_1 \in \mathbb{R}^{M\times N_1}$, $A_2 \in \mathbb{R}^{M\times N_2}$, and $C \in \mathbb{R}^M$. The Lagrangian of this problem is $L(x_1,x_2,\lambda) = f_1(x_1) + f_2(x_2) + \lambda^T(A_1x_1 + A_2x_2 - C)$, where $\lambda \in \mathbb{R}^M$ is the dual variable. Define $\mathcal{X}_1^* \times \mathcal{X}_2^* \times \Lambda^*$ as the set of saddle-points of $L(x_1,x_2,\lambda)$ on $\mathcal{X} \times \Omega \times \mathbb{R}^M$. It assumes that f_1 is μ_{f_1}-strongly convex, twice differentiable and smooth, f_2 is Lipschitz continuous and strongly convex on Ω, and A_1 has full row rank. To solve the problem (14), it provides the following PDDS.

$$x_1(t) = \arg\min_{s\in\mathcal{X}}\{f_1(s) + (\lambda(t))^T A_1 s\}$$

$$\dot{x}_2(t) \in \mathcal{P}_{\mathcal{T}_\Omega(x_2(t))}(-\partial f_2(x_2(t)) - A_2^T\lambda(t))$$

$$\dot{\lambda}(t) = \nabla\Phi(\lambda(t)) + A_2x_2(t) - C \tag{15}$$

where $\mathcal{T}_\Omega(x_2(t))$ is the tangent cone to Ω at $x_2(t)$, and $\Phi(\lambda(t)) = \min_{s\in\mathcal{X}}\{f_1(s) + (\lambda(t))^T A_1 s\}$. In [8, Theorem 7], it shows that the trajectory $(x_2(t),\lambda(t))$ associated with PDDS (15) is exponentially convergent under the above assumptions, without giving the convergence rate of $x_1(t)$. Via similar analysis technology with Theorem 2, the convergence properties of the PDDS (15) is explored without the assumption that A_1 has full row rank.

Corollary 1: Supposing that f_1 is μ_{f_1}-strongly convex and smooth, f_2 is Lipschitz continuous and strongly convex on Ω. Given $x_2(0) \in \Omega$, $\lambda(0) \in \mathbb{R}^M$, the solution of PDDS (15) converges exponentially to $\mathcal{X}_2^* \times \Lambda^*$, i.e., there exist ν_1, $\nu_2 > 0$ such that

$$\|x_2(t) - x_2^*\| + \|\lambda(t) - [\lambda]^*\| \leq \nu_1 e^{-\nu_2 t}, \quad t \geq 0.$$

Moreover, there exist $\nu_2, \nu_3 > 0$ such that

$$\Phi([\lambda]^*) - \Phi(\lambda) \leq \nu_3 \cdot e^{-\nu_2 t}, \quad t \geq 0,$$

$$\|x_1(t) - x_1^*\| \leq \sqrt{\frac{4\nu_3}{\mu_{f_1}}} \cdot e^{-\frac{\nu_2}{2}t}, \quad t \geq 0.$$

Proof: See Appendix C. ■

Next, we reanalyze the convergence of the fixed-time convergent PDDS provided in [9] without the assumption that A_1 has full row rank. Consider the following dynamical system:

$$\dot{\lambda}(t) = -c_1\frac{\nabla g(\lambda)}{\|\nabla g(\lambda)\|^{\frac{p_1-2}{p_1-1}}} - c_2\frac{\nabla g(\lambda)}{\|\nabla g(\lambda)\|^{\frac{p_2-2}{p_2-1}}} \tag{16}$$

where $g(\lambda) = -d(\lambda)$, $c_1, c_2 > 0$, $p_1 > 2$, and $1 < p_2 < 2$. Similar as the proof of [9] and Theorem 2, we have the following corollary.

Corollary 2: Supposing that Assumption 1 holds, the trajectory $\lambda(t)$ of dynamics (16) converges to Λ^* within a fixed time for any initial condition $\lambda(0) \in \mathbb{R}^m$. Moreover, there exists $T_\lambda > 0$ such that

$$\lim_{t\to T_\lambda} f_{opt} - d(\lambda(t)) \to 0, \ \lim_{t\to T_\lambda} \|A_1x(t) - b\| \to 0$$

$$\lim_{t\to T_\lambda} \|x(t) - x^*\| \to 0, \ \lim_{t\to T_\lambda} |f(x(t)) - f_{opt}| \to 0.$$

Furthermore, it holds that $T_\lambda \leq 4/k_1(2-\alpha_1) + 4/k_2(\alpha_2 - 2)$, where $\alpha_1 = 2-(p_1-2)/(p_1-1)$ and $\alpha_2 = 2-(p_2-2)/(p_2-1)$, $k_1 = c_1 2^{(2+\alpha_1)/4}(2l_d l_f^2\omega^2)^{-(\alpha_1/2)}$ and $k_2 = c_2 2^{(2+\alpha_2)/4}(2l_d l_f^2\omega^2)^{-(\alpha_2/2)}$.

Illustrative examples:

In this subsection, we conduct two numerical experiments to corroborate our theoretical analysis. Firstly, consider the COP: $\min_{x\in\mathbb{R}^n}(1/2)\|\mathbf{U}x - \mathbf{V}\|^2$, s.t. $\mathbf{A}x = \mathbf{b}$, $\mathbf{C}x \geq \mathbf{0}$, where $\mathbf{U} \in \mathbb{R}^{2000\times 100}$, $\mathbf{V} \in \mathbb{R}^{100}$, $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{100\times 100}$, and $\mathbf{b} \in \mathbb{R}^{100}$, which are generated randomly. The numerical results are shown in the left of Fig. 1, which illustrates the exponential convergence of PDDS (3).

Then, consider the distributed optimization over the undirected and connected network with N computational agents: $x^* \in \arg\min_{x\in\mathbb{R}^{20}} \sum_{i=1}^N (1/2)\|U_i x - V_i\|^2$, where $U_i \in \mathbb{R}^{20\times 20}$ is the measurement matrix generated randomly such that $U_i^T U_i$ is positive definite, and $V_i \in \mathbb{R}^{20}$ generated randomly is the noise vector. Here, we set the communication topology as a line graph with 4 nodes. By [16], the problem is equivalent to $\min\ F(x) = \sum_{i=1}^4 (1/2)\|U_i x^i - V_i\|^2$, s.t.

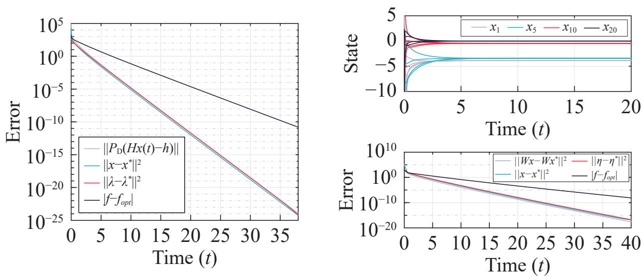

Fig. 1. The numerical results of considered examples.

$-\sqrt{\mathcal{L}\otimes I_{20}}x=0$, where $\mathcal{L}$ is the Laplacian matrix of the considered network, $x^i\in\mathbb{R}^{20}$ is the local variable of agent i, and $x=[(x^1)^T,\ldots,(x^4)^T]^T$. Let $\boldsymbol{\eta}=\sqrt{\mathcal{L}\otimes I_{20}\boldsymbol{\lambda}}$, where $\boldsymbol{\lambda}=[(\lambda^1)^T,\ldots,(\lambda^4)^T]^T$, and λ^i is the local dual variable. Then the PDDS (8) for this problem is

$$x^i(t)=\arg\min_s\{f_i(s)-\eta^i(t)^T s\},\ \ i=1,\ldots,4$$

$$\dot{\eta}^i(t)=-\sum_{j=1}^{4}\mathcal{L}_{ij}\cdot\nabla f_i^*(\eta^i(t)),\ \ i=1,\ldots,4 \tag{17}$$

where f_i^* is the conjugate function of $(1/2)\|U_ix-V_i\|^2$. The state transient behaviors of the 4 agents and the performance of the algorithm (17) are shown as in the right of Fig. 1.

Conclusions: In this letter, we have studied the convergence properties of PDDS (3) and (8) to solve COPs with linear constraints. Under weaker assumptions, we have shown the exponential convergence of (3) and (8). In addition, applying the analysis technique in Theorems 1 and 2, the convergence results in [8] and [9] have been improved.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (61833005, 62003084), and the Natural Science Foundation of Jiangsu Province of China (BK20200355).

Appendix A:

Proof of Theorem 1: Consider a Lyapunov function as follows, $V_1(\lambda)=\upsilon g(\lambda)-\upsilon g([\lambda]^*)-\upsilon(\lambda-[\lambda]^*)^T\nabla g([\lambda]^*)+(1/2)\|\lambda-[\lambda]^*\|^2$, where $g(\lambda)=-d(\lambda)$. Since $g(\lambda)$ is convex, one has $g(\lambda)-g([\lambda]^*)-(\lambda-[\lambda]^*)^T\nabla g([\lambda]^*)\geq 0$, which implies that $V_1(\lambda)\geq(1/2)\|\lambda-[\lambda]^*\|^2$. On the other hand, by the l_d-smoothness of $g(\lambda)$, it holds that $g(\lambda)-g([\lambda]^*)-(\lambda-[\lambda]^*)^T\nabla g([\lambda]^*)\leq(l_d/2)\|\lambda-[\lambda]^*\|^2$. Thus, it deduces that $(1/2)\|\lambda-[\lambda]^*\|^2\leq V_1(\lambda)\leq((\upsilon l_d+1)/2)\|\lambda-[\lambda]^*\|^2$. Differentiating $V_1(\lambda)$ along the solutions of (3), one gets

$$\frac{d}{dt}V_1(\lambda)=-(\lambda-\upsilon\nabla g(\lambda)-\mathcal{P})^T(\mathcal{P}-[\lambda]^*)-\|\lambda-\mathcal{P}\|^2$$
$$-\upsilon(\mathcal{P}-[\lambda]^*)^T\nabla g([\lambda]^*)-\upsilon(\nabla g(\lambda)-\nabla g([\lambda]^*))^T(\lambda-[\lambda]^*)$$

where $\mathcal{P}=\mathcal{P}(\lambda+\upsilon\cdot\nabla d(\lambda))$. By the basic property of projection operator, one has $(\lambda-\upsilon\nabla g(\lambda)-\mathcal{P})^T(\mathcal{P}-[\lambda]^*)\geq 0$. Since $[\lambda]^*\in\Lambda^*$, it holds that $(\mathcal{P}-[\lambda]^*)^T\nabla g([\lambda]^*)\geq 0$. Moreover, the convexity of $g(\lambda)$ implies that $(\nabla g(\lambda)-\nabla g([\lambda]^*))^T(\lambda-[\lambda]^*)\geq 0$. Under Assumption 1 and strong Slater conditions, by [18, Theorem 3.2], it gives that the following global error bound like property:

$$\kappa\cdot\|\lambda-[\lambda]^*\|^2\leq\left\|\lambda-\mathcal{P}_{\mathbb{D}}\left(\lambda+\upsilon\cdot\nabla d(\lambda)\right)\right\|^2,\ \ \forall\lambda\in\mathbb{D} \tag{18}$$

where $\kappa>0$ is a constant independent on λ. Hence, it holds that

$$\frac{d}{dt}V_1(\lambda)\leq-\|\lambda-\mathcal{P}\|^2\leq-\kappa\|\lambda-[\lambda]^*\|^2\leq-2\beta V_1(\lambda)$$

where $\beta=\kappa/(\upsilon l_d+1)$. Then, it holds that $V_1(\lambda)\leq V_1(\lambda(0))\cdot\exp(-\beta t),\forall t\geq 0$. By $\frac{1}{2}\|\lambda-[\lambda]^*\|^2\leq V_1(\lambda)$, it has $\mathrm{dist}(\lambda(t),\Lambda^*)\leq\sqrt{2V_1(\lambda(0))}\cdot\exp(-\beta t),\quad\forall t\geq 0$. It implies that $\lambda(t)$ converges to Λ^* exponentially, and convergence rate is no less than β. In the next, we fully consider the convergence properties of PDDS (3).

Step 1: Show the exponential convergence of $f_{opt}-d(\lambda(t))$. Since $-d(\lambda)$ is l_d-smooth, it has that $g(\lambda)-g([\lambda]^*)-(\lambda-[\lambda]^*)^T\nabla g([\lambda]^*)\leq(l_d/2)\|\lambda-[\lambda]^*\|^2$. Thus,

$$f_{opt}-d(\lambda)\leq(\lambda-[\lambda]^*)^T\nabla g([\lambda]^*)+\frac{l_d}{2}\|\lambda-[\lambda]^*\|^2$$
$$\leq M_1\|\lambda-[\lambda]^*\|+\frac{l_d}{2}\|\lambda-[\lambda]^*\|^2\leq\alpha\cdot e^{-\beta t},\quad\forall t\geq 0$$

where $\alpha=2\max\{M_1\sqrt{2V_1(\lambda(0))},\ V_1(\lambda(0))l_d\}$, $M_1=\max_{\lambda\in\Lambda^*}\|\nabla g(\lambda)\|$.

Step 2: Show the exponential convergence of $\|\mathcal{P}_{\mathbb{D}}(Hx(t)-h)\|$. By the basic property of projection operator, it gives that $(\lambda-\mathcal{P})^T(\lambda-\mathcal{P})\leq\upsilon\nabla g(\lambda)^T(\lambda-\mathcal{P})$. Thus, it holds that

$$\|\mathcal{P}-\lambda\|^2\leq 2\upsilon\nabla g(\lambda)^T(\lambda-\mathcal{P})-\|\mathcal{P}-\lambda\|^2$$
$$=(l_d\upsilon-1)\|\mathcal{P}-\lambda\|^2-2\upsilon\Big(\nabla g(\lambda)^T(\mathcal{P}-\lambda)+\frac{l_d}{2}\|\mathcal{P}-\lambda\|^2\Big).$$

Since $\upsilon\leq 1/l_d$, and $g(\lambda)$ is l_d-smooth, it gives that $l_d\upsilon-1\leq 0$ and $g(\mathcal{P})-g(\lambda)-(\mathcal{P}-\lambda)^T\nabla g(\lambda)\leq(l_d/2)\|\lambda-\mathcal{P}\|^2$. Then, it holds that $\|\mathcal{P}-\lambda\|^2\leq 2\upsilon(g(\lambda)-g(\mathcal{P}))\leq 2\upsilon(f_{opt}-d(\lambda)$.

Then, to prove the exponential convergence of $\|\mathcal{P}_{\mathbb{D}}(Hx(t)-h)\|$, it is sufficient to show that $\upsilon\|\mathcal{P}_{\mathbb{D}}(\nabla d(\lambda(t)))\|^2\leq(1/\upsilon)\|\lambda(t)-\mathcal{P}\|^2$. Inspired by [18], we will prove this inequality componentwise. First, recalling that $\mathbb{D}=\mathbb{R}^{m_1}\times\mathbb{R}_+^{m_2}$, for $i=1,\ldots,m_1$, it gets that $\upsilon\cdot\nabla_i d(\lambda)=\mathcal{P}_i-\lambda_i$. Thus, it holds that $\upsilon\cdot|\mathcal{P}_{\mathbb{R}}(\nabla_i d(\lambda(t)))|^2=\upsilon\cdot|\nabla_i d(\lambda(t))|^2=(1/\upsilon)|\lambda_i-\mathcal{P}_i|^2$. Then, set the disjoint sets as

$$I_-=\{i\in[m_1+1,m_1+m_2]:\nabla_i d(\lambda)<0\}$$
$$I_+=\{i\in[m_1+1,m_1+m_2]:\nabla_i d(\lambda)\geq 0\}.$$

It gives that $\upsilon\cdot|\mathcal{P}_{\mathbb{R}_+}(\nabla_i d(\lambda))|^2=0\leq(1/\upsilon)|\lambda_i-\mathcal{P}_i|^2,\quad\forall i\in I_-$. On the other hand, for any $i\in I_+$, one has $\upsilon\cdot|\mathcal{P}_{\mathbb{R}_+}(\nabla_i d(\lambda))|^2=\upsilon\cdot|\nabla_i d(\lambda)|^2=(1/\upsilon)|\lambda_i-\mathcal{P}_i|^2$. In summary, it obtains

$$\upsilon\|\mathcal{P}_{\mathbb{D}}(Hx(t)-h)\|^2\leq\frac{1}{\upsilon}\|\lambda(t)-\mathcal{P}\|^2\leq\frac{1}{\upsilon}\cdot 2\upsilon(f_{opt}-d(\lambda).$$

Then, (5) immediately holds.

Step 3: Show the exponential convergence of $\|x(t)-x^*\|$. It follows from f is μ_f-strongly convex that $L(x,\lambda(t))=f(x)+\lambda(t)^T(Hx-h)$ is μ_f-strongly convex in the variable x. Then, it deduces that

$$\frac{\mu_f}{2}\|x(t)-x^*\|^2\leq L(x^*,\lambda(t))-L(x(t),\lambda(t)).$$

On the other hand, $x(t)=\arg\min_{s\in\mathbb{R}^n}L(s,\lambda(t))$ implies that $d(\lambda(t))=f(x(t))+\lambda(t)^T(Hx(t)-h)$ and $\nabla d(\lambda(t))=Hx(t)-h$. Thus, it concludes that $L(x^*,\lambda(t))-L(x(t),\lambda(t))=f_{opt}-d(\lambda)+\lambda(t)^T(Hx^*-h)$. Since $A_1x^*=b_1$, and $A_2x^*\leq b_2$, one has $\lambda(t)^T(Hx^*-h)=\varphi(t)^T(A_2x^*-b_2)\leq 0$. Hence, (6) holds.

Step 4: Show the exponential convergence of $|f(t)-f_{opt}|$. For any $\lambda^*\in\Lambda^*$, it follows from $x(t)=\arg\min_{s\in\mathbb{R}^n}L(s,\lambda(t))$ that

$$f_{opt}=f(x(t))+(\phi^*)^T(A_1x-b_1)+(\varphi^*)^T(A_2x-b_2).$$

Since $\varphi^*\in\mathbb{R}_+^{m_2}$, it can be shown that $(\lambda^*)^T(Hx(t)-h)\leq(\lambda^*)^T\mathcal{P}_{\mathbb{D}}(Hx(t)-h)$. Then, it holds that $f_{opt}-f(x(t))\leq M_2\|\mathcal{P}_{\mathbb{D}}(Hx(t)-h)\|$, where $M_2=\max_{\lambda\in\Lambda^*}\|\lambda\|$. Since $f(x)-f(x^*)-(x-x^*)^T\nabla f(x^*)\leq(l_f/2)\|x-x^*\|^2$, and $x^*=\arg\min f(x)+(\lambda^*)^T(Hx-h)$, it holds that $\nabla f(x^*)=H^T\lambda^*$, which implies that $f(x(t))-f_{opt}\leq M_3\|x(t)-x^*\|+(l_f/2)\|x(t)-x^*\|^2$, where $M_3=\max_{\lambda\in\Lambda^*}\|H^T\lambda\|$. Therefore, it holds that $|f(x(t))-f_{opt}|\leq\gamma\exp(-\beta t/2),\ t\geq 0$, where $\gamma=3\cdot\max\{M_2\sqrt{2\alpha/\upsilon},M_3\sqrt{\{2\alpha/\mu_f\}},\alpha l_f/\mu_f\}$. ■

Appendix B:

Proof of Theorem 2: First, we show the exponential convergence of dual variable. Consider the Lyapunov function $E(\lambda)=(1/2)\|\lambda-[\lambda]^*\|^2$. Since $g(\lambda)=-d(\lambda)$, differentiating $E(\lambda)$ along the solution of (8b) yields $\frac{d}{dt}E(\lambda)=-(\lambda-[\lambda]^*)^T\nabla g(\lambda)\leq-(1/2l_f\omega)\|\lambda-$

$[\lambda]^*\|^2$. The first inequality holds due to (9). Thus, it holds that $\frac{d}{dt}E(\lambda) \leq -(1/l_f\omega)E(\lambda)$. Then, one has $\text{dist}(\lambda(t),\Lambda^*) \leq \bar{\alpha}\cdot\exp(-\bar{\beta}t/2)$, $t \geq 0$, where $\bar{\alpha} = \|\lambda(0)-[\lambda]^*\|$, and $\bar{\beta} = 1/l_f\omega$. It implies that $\lambda(t)$ converges to Λ^* exponentially.

Step 1: Show the exponential convergence of $f_{opt} - d(\lambda(t))$. Consider another Lyapunov function $G(\lambda) = (1/2)(g(\lambda) - g_{opt})$, where $g_{opt} = -f_{opt}$. Differentiating $G(\lambda)$ along the solution of (8b) yields $\frac{d}{dt}G(\lambda) = -(g - g_{opt})\cdot\|\nabla g\|^2$. It follows from (9) that $(\lambda - [\lambda]^*)^T\nabla g(\lambda) \geq 1/(2l_f\omega)\|\lambda - [\lambda]^*\|^2$. By Cauchy-Schwartz, one has $\|\nabla g(\lambda)\| \geq (1/2l_f\omega)\|\lambda - [\lambda]^*\|$. Then, by the smoothness of g and $\nabla g([\lambda]^*) = 0$, one has $g(\lambda) - g_{opt} \leq 2l_d l_f^2\omega^2\cdot\|\nabla g(\lambda)\|^2$. Therefore, it holds that $\frac{d}{dt}G(\lambda) \leq -((g - g_{opt})^2/2l_d l_f^2\omega^2) = -(G(\lambda)/l_d l_f^2\omega^2)$. Thus, there exist $\tilde{\alpha} > 0$ and $\tilde{\beta} > 0$ such that $f_{opt} - d(\lambda(t)) \leq \tilde{\alpha}\cdot\exp(-\tilde{\beta}t)$, $\forall t \geq 0$, where $\tilde{\alpha} = g(0) - g_{opt}$, and $\tilde{\beta} = (2l_d l_f^2\omega^2)^{-1}$.

Step 2: Show the exponential convergence of $\|A_1x(t) - b_1\|$. It follows from $\nabla d(\lambda) = A_1x(t) - b_1$ that $\|A_1x(t) - A_1x^*\| = \|\nabla g(\lambda) - \nabla g([\lambda]^*)\| \leq l_d\|\lambda(t) - [\lambda]^*\| \leq l_d\bar{\alpha}\cdot\exp(-\bar{\beta}t/2)$, $t \geq 0$. The first inequality holds due to the smoothness of $g(\lambda)$, and the second inequality holds due to the exponential convergence of dual variable. Thus, $A_1x(t) - b_1$ is exponentially convergent.

Step 3: Show the exponential convergence of $\|x(t) - x^*\|$. Similar as the proof of Theorem 3.1 – Step 3, it can be proved that $(\mu_f/2)\|x(t) - x^*\|^2 \leq f_{opt} - d(\lambda)$. Hence, it holds that $(\mu_f/2)\|x(t) - x^*\|^2 \leq f_{opt} - d(\lambda) \leq \tilde{\alpha}\cdot\exp(-\tilde{\beta}t/2)$, $\forall t > 0$.

Step 4: Show the exponential convergence of $|f(t) - f_{opt}|$. Similar as the proof of Theorem 3.1 – Step 4, it can be proved that for any $\lambda^* \in \Lambda^*$, it follows from $x(t) = \arg\min_{s\in\mathbb{R}^n} L(s,\lambda(t))$ that $f_{opt} - f(x(t)) \leq \|s^*\|\cdot\|x(t) - x^*\|$, $\forall t \geq 0$. Since $\nabla f(x^*) = A_1^T\lambda^* = s^*$,whichimpliesthat$f(x(t)) - f_{opt} \leq \|s^*\|\cdot\|x(t) - x^*\| + (l_f/2)\|x(t) - x^*\|^2$. Therefore, it holds that $|f(x(t)) - f_{opt}| \leq \tilde{\gamma}\exp(-\tilde{\beta}t/2)$, $t \geq 0$, where $\tilde{\gamma} = 2\cdot\max\{\|s^*\|\sqrt{2\tilde{\alpha}/\mu_f}, l_f\tilde{\alpha}/\mu_f\}$. ■

Appendix C:

Proof of Corollary 1: First, we show the exponential convergence of PDDS (15) without the assumption that A_1 has full row rank. Since f_1 and f_2 are strongly convex, the problem (14) has a unique solution (x_1^*, x_2^*). Consider the following equality constraints problem.

$$\min_{x_1\in\mathcal{X}} f_1(x_1) + f_2(x_2^*), \quad \text{s.t.} \quad A_1x_1 = C - A_2x_2^*. \tag{19}$$

The Lagrangian of this problem is $L(x_1, x_2^*, \lambda) = f_1(x_1) + f_2(x_2^*) + \lambda^T(A_1x_1 + A_2x_2^* - C)$. It holds that the optimal solution of problem (19) is x_1^*. Moreover, problem (14) and problem (19) have the same dual optimal solution set Λ^*. Define the dual function of (19) as $\tilde{\Phi}(\lambda) = -f_1^*(-A_1^T\lambda) - (C - A_2x_2^*)^T\lambda$, where f_1^* is the convex conjugate of f_1. Because f_1 is strongly convex and smooth, the function f_1^* is strongly convex and smooth. Thus, by [13, Theorem 9], it holds that there exists a constant $\tilde{\kappa} > 0$ only related to A_1 such that $(\lambda - [\lambda]^*)^T\nabla(-\tilde{\Phi})(\lambda) \geq \tilde{\kappa}\|\lambda - [\lambda]^*\|^2$, where $[\lambda]^*$ is the projection of λ onto Λ^*, and $\nabla(-\tilde{\Phi}(\lambda)) = -A_1(\nabla f_1^*)(-A_1^T\lambda) + (C - A_2x_2^*)$. By KKT condition, one has $A_1x_1^* + A_2x_2^* - C = 0$. Thus, $\nabla(-\tilde{\Phi}(\lambda)) = \nabla(-\Phi(\lambda)) + A_1x_1^*$. Then, we can prove that $\nabla(-\tilde{\Phi}(\lambda(t))) = \nabla(-\Phi(\lambda(t))) - \nabla(-\Phi([\lambda]^*))$, which implies that $(\lambda(t) - [\lambda]^*)^T \times (\nabla(-\Phi)(\lambda(t)) - \nabla(-\Phi)([\lambda]^*)) \geq \tilde{\kappa}\|\lambda - [\lambda]^*\|^2$.

Then, define the Lyapunov function candidate $V(x_2,\lambda) = (1/2)\|x_2 - x_2^*\|^2 + (1/2)\|\lambda - [\lambda]^*\|$, similar as the proof of [8, Theorem 7], it can be shown that there exists $\nu_1, \nu_2 > 0$ such that

$$\|x_2(t) - x_2^*\| + \|\lambda(t) - [\lambda]^*\| \leq \nu_1 e^{-\nu_2 t}, \quad t \geq 0.$$

It implies that $(x_2(t), \lambda(t))$ converges to $\mathcal{X}_2^* \times \Lambda^*$ exponentially. In the next, similar as the proof of Theorem 2, we consider other convergence properties of PDDS (15). Then, similar as the proof of Theorem (2), we can complete the rest of the proof. ■

References

[1] W. Lin, Y. Wang, C. Li, and X. Yu, "Distributed resource allocation via accelerated saddle point dynamics," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 9, pp. 1588–1599, 2021.

[2] C. Xu and X. He, "A Fully distributed approach to optimal energy scheduling of users and generators considering a novel combined neurodynamic algorithm in smart grid," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 7, pp. 1325–1335, 2021.

[3] Z. Liu, Y. Yuan, X. Guan and X. Li, "An approach of distributed joint optimization for cluster-based wireless sensor networks," *IEEE/CAA Journal of Automatica Sinica*, vol. 2, no. 3, pp. 267–273, 2015.

[4] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *J. Nonlinear Sci.*, vol. 29, pp. 1247–1272, 2019.

[5] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 43–48, 2019.

[6] Y. Tang, G. Qu, and N. Li, "Semi-global exponential stability of augmented primal-dual gradient dynamics for constrained convex optimization," *Syst. Control Lett.*, vol. 144, Article No. 104754, 2020.

[7] X. Chen and N. Li, "Exponential stability of primal-dual gradient dynamics with non-strong convexity, " in *Proc. American Control Conf., IEEE*, 2020, pp. 1612–1618.

[8] Z. Wang, W. Wei, C. Zhao, Z. Ma, Z. Zheng, Y. Zhang, and F. Liu, "Exponential stability of partial primal-dual gradient dynamics with nonsmooth objective functions," *Automatica*, vol. 129, Article No. 109585, 2021.

[9] K. Garg and D. Panagou, "Fixed-time stable gradient flows: Applications to continuous-time optimization," *IEEE Trans. Autom. Control*, vol. 66, no. 5, pp. 2002–2015, 2021.

[10] S. A. Alghunaim and A. H. Sayed, "Linear convergence of primal-dual gradient methods and their performance in distributed optimization," *Automatica*, vol. 117, Article No. 109003, 2020.

[11] W. Bian and X. Xue, "Asymptotic behavior analysis on multivalued evolution inclusion with projection in Hilbert space," *Optimization*, vol. 64, no. 4, pp. 853–875, 2015.

[12] N. T. T. Ha, J. J. Strodiot, and P. T. Vuong, "On the global exponential stability of a projected dynamical system for strongly pseudomonotone variational inequalities," *Optim. Lett.*, vol. 12, pp. 1625–1638, 2018.

[13] I. Necoara, Y. Nesterov, and F. Glineur, "Linear convergence of first order methods for non-strongly convex optimization," *Math. Program.*, vol. 175, pp. 69–107, 2019.

[14] C. Shi and G. Yang, "Augmented lagrange algorithms for distributed optimization over multi-agent networks via edge-based method," *Automatica*, vol. 94, pp. 55–62, 2018.

[15] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.

[16] Q. Liu and J. Wang, "A second-order multi-agent network for boundconstrained distributed optimization," *IEEE Trans. Autom. Control*, vol. 60, no. 12, pp. 3310–3315, 2015.

[17] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, 2014.

[18] I. Necoara and V. Nedelcu, "On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems," *Automatica*, vol. 5, no. 5, pp. 209–216, 2015.