ACCELERATED PRIMAL-DUAL MIRROR DYNAMICAL APPROACHES FOR CONSTRAINED CONVEX OPTIMIZATION *

YOU ZHAO †, XIAOFENG LIAO ‡, XING HE §, AND CHAOJIE LI, ¶

Abstract. This paper investigates two accelerated primal-dual mirror dynamical approaches for smooth and nonsmooth convex optimization problems with affine and closed, convex set constraints. In the smooth case, an accelerated primal-dual mirror dynamical approach (APDMD) based on accelerated mirror descent and primal-dual framework is proposed and accelerated convergence properties of primal-dual gap, feasibility measure and the objective function value along with trajectories of APDMD are derived by the Lyapunov analysis method. Then, we extend APDMD into two distributed dynamical approaches to deal with two types of distributed smooth optimization problems, i.e., distributed constrained consensus problem (DCCP) and distributed extended monotropic optimization (DEMO) with accelerated convergence guarantees. Moreover, in the nonsmooth case, we propose a smoothing accelerated primal-dual mirror dynamical approach (SAPDMD) with the help of smoothing approximation technique and the above APDMD. We further also prove that primal-dual gap, objective function value and feasibility measure of SAPDMD have the same accelerated convergence properties as APDMD by choosing the appropriate smooth approximation parameters. Later, we propose two smoothing accelerated distributed dynamical approaches to deal with nonsmooth DEMO and DCCP to obtain accelerated and efficient solutions. Finally, numerical experiments are given to demonstrate the effectiveness of the proposed accelerated mirror dynamical approaches.

Key words. accelerated mirror dynamical approaches, smooth and nonsmooth optimization, constrained optimization, smoothing approximation, distributed approaches

AMS subject classifications. 90C25, 90C06

1. Introduction.

1.1. Problem statement. In this paper, we consider the following convex constrained optimization problem:

\[ \min_{x \in \mathbb{R}^n} f(x), \text{ s.t. } Ax = b, x \in X, \]

where

\[ \begin{align*}
\text{(1.2)} & \\
\text{X is a closed and convex set,} & \\
f : \mathcal{X} \to \mathbb{R} \text{ may be a smooth or nonsmooth convex function,} & \\
A : \mathcal{X} \to \mathbb{R}^m \text{ is a continuous linear operator and } b \in \mathbb{R}^m, & \\
\text{The optimal solution set of problem (1.1) is nonempty.} & 
\end{align*} \]

* Funding: This work was supported in part by the National Key R&D Program of China (No. 2018AAA0100101), in part by National Natural Science Foundation of China (Grant no. 61932006, 61772434), in part by Chongqing Technique Innovation and Application Development Project (Grant no. cstc2020jxscx-msxzmX0156) and in part by the Fundamental Research Funds for the Central Universities (Project No. XDJK2020(TY003)).

† College of Computer Science, Chongqing University, Chongqing 400044, PR. China. (Zhaoyou1991sdztz@163.com).

‡ College of Computer Science, Chongqing University, Chongqing 400044, PR. China. (xfliao@cqu.edu.cn). (Corresponding author: Xiaofeng Liao)

§ Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, School of Electronics and Information Engineering, Southwest University, Chongqing 400715, China. (hexingdoc@swu.edu.cn).

¶ School of Electrical Engineering and Telecommunications, University of New South Wales, Kensington, NSW 2052, Australia. (chaojie.li@unsw.edu.au).

arXiv:2205.15983v1 [math.OC] 31 May 2022
This problem covers many optimization problems in various applied fields such as machine learning, sparse signal reconstruction, image deblurring, resource allocation, saddle point, and network optimization problems (see, e.g. [16, 29, 37, 42, 51]).

1.2. Historical presentation. The inertial (secend-order or accelerated) dynamical approaches are increasingly popular for solving the unconstrained optimization problem

$$\min f(x).$$

To deal with problem (1.3), Polyak [44] first proposed the heavy ball with friction dynamical approach

$$\ddot{x} + \eta \dot{x} + \nabla f(x) = 0,$$

where $\eta > 0$ is a damping parameter. Later, in [2], Alvarez studied the asymptotic behavior of (1.4) with a time independent parameter $\eta$ when $f(x)$ is convex. Bégout et al. [12] studied some convergence properties of the HBF (1.4) with a constant $\eta$ when $f(x)$ is a nonconvex function. Aujol et al. [6, 7] investigated the convergence rate of HBF dynamical system and its corresponding discrete algorithm with a fixed parameter $\eta$ under the condition that $f(x)$ satisfies quasi-strongly convex and Lojasiewicz properties, respectively. Su, Boyd, and Candès [47] first revealed that the HBF (1.4) with $\eta = \alpha t$, $\alpha \geq 3$ can be regarded as the continuous-time limit of the Nesterov’s accelerated gradient algorithm and it has an accelerated convergence rate, i.e., $f(x(t)) - \min f(x) = O\left(\frac{1}{t^2}\right)$. May [23] further improved the convergence rate of it to $f(x(t)) - \min f(x) = o\left(\frac{1}{t^2}\right)$. When $\eta = \frac{\alpha}{t^2}$, $0 < \alpha \leq 3$, the convergence rate $f(x(t)) - \min f(x) = O\left(t^{-\frac{2}{3}}\right)$ was estimated in [5] and [3] for smooth and nonsmooth convex functions $f(x)$. Cabot and Frankel [19] studied the asymptotic behavior of the HBF (1.4) when $\eta = \alpha t^\gamma$, $0 < \gamma < 1$. Moreover, Wibisono et al. [48] studied a series of accelerated (inertial) dynamical approaches to the problem (1.3) by the Bregman Lagrangian based on the calculus of variations. By the manifold with curvature bounded from below, Alimisis et al. [1] proposed a Riemannian variant of accelerated gradient dynamical approach.

Recently, to solve the problem (1.3) with a closed and convex set constraint $\mathcal{X}$, i.e.,

$$\min f(x), x \in \mathcal{X} \subset \mathbb{R}^n,$$

many inertial dynamical approaches have been studied. Boţ et al. [17] proposed an inertial forward-backward dynamical approach. Based on the projection operators, He et al. [28] proposed an inertial dynamical approach to solve the problem (1.5) with non-convex objective functions. Later, based on the work in [28] and smoothing approximation technique, two smoothing inertial hydrodynamic approaches were proposed to solve constrained nonconvex $L_{p-q}$ minimization problem to reconstruct the sparse signal in [53]. Combining the dynamical approach of mirror descent with the continuous version of Nesterov’s acceleration algorithm, Krichene et al. [36] proposed an accelerated mirror dynamical approach for problem (1.5) as follows:

$$\begin{align*}
\dot{X} &= \frac{\gamma}{t} (\nabla \psi^* (Z) - X), \\
\dot{Z} &= -\frac{1}{\gamma} \nabla f(X), \\
X(0) &= x_0, Z(0) = z_0, \text{ with } \nabla \psi^*(z_0) = x_0,
\end{align*}$$
where $\gamma \geq 2$ and $\nabla \psi^*$ is the gradient of conjugate of $\psi$ (see Subsection 2.3).

In addition, in order to solve the problem (1.3) with an affine constraint, i.e,
\[
\min f(x), Ax = b,
\]
many inertial dynamical approaches under primal-dual framework were extensively investigated. Zeng et al. [50] first proposed a second-order dynamical primal-dual approach for solving the problem (1.7) and proved that the convergence rate of the gap between the Lagrangian function and its optimal value is $O\left(\frac{1}{t^2}\right)$. The corresponding inertial dynamical approach is then extended to solve the distributed optimization problems. Later, Bot et al. [18] improved the convergence rates of the works in [50] and provided the weakly convergence analysis to a primal-dual optimal solution of the problem (1.7). He et al. [30] studied an inertial primal-dual dynamical approach without/with perturbations for separable convex optimization problems with an affine constraint. Attouch et al. [4] proposed a temporally rescaled inertial augmented Lagrangian system (TRIALS) with three time-varying parameters (i.e., viscous damping, extrapolation and temporal scaling) to address separable smooth/nonsmooth convex optimization problems with an affine constraint, and presented the fast convergence properties of TRIALS. In addition, He et al. [31] and Luo et al. [39] further proposed a “second-order”+“first-order” primal-dual dynamical approaches for solving problem (1.7) with accelerated convergence guarantees.

However, many practical problems in machine learning, sparse signal reconstruction, image deblurring, resource allocation and so on, not only have affine constraints, but also have set constraints, as in problem (1.1). The problem (1.1) can be regarded as a generalization for the problems (1.3), (1.5), (1.7), (3.1) and (3.2). In recent years, some first order dynamical approaches [38, 45, 49, 51] based on primal-dual framework and projection were proposed to solve the problem (1.1) only with convergence analysis or ergodic $O\left(\frac{1}{t}\right)$ convergence rate. To the best of our knowledge, accelerated (inertial) dynamical approaches based on primal-dual framework with non-ergodic $O\left(\frac{1}{t^2}\right)$ convergence rates are rarely involved. It is worth noting that the inertial dynamical approaches mentioned above were used for solving the problems (1.3), (1.5) and (1.7) and they cannot be used directly for the problem (1.1).

This paper aims to investigate accelerated primal-dual dynamical approaches based on mirror descent and smoothing approximation methods for solving problem (1.1) with an accelerated non-ergodic convergence rate $O\left(\frac{1}{t^2}\right)$. Our contributions are summarized as follows:

- For the problem (1.1) in the smooth case, an accelerated primal-dual mirror dynamical approach (APDMD) is proposed for the first time, it is used to solve problem (1.1) with accelerated convergence rates of Lagrangian and objective functions’ gaps. We provide an interpretation of the APDMD from different perspectives (i.e., neurodynamic approach, Hamilton’s system, game theory and control theory). Moreover, based on the properties of conjugate function and Cauchy-Lipschitz-Picard theorem, the feasibility (i.e, ensuring that the trajectories of solutions always satisfy the set constraint), existence and uniqueness of the global solution of APDMD are obtained. Last, applying the APDMD to DCCP and EDMO leads to two distributed APDMDs (ADPDMD and ADMD) with accelerated convergence guarantees.

- For the problem (1.1) in the nonsmooth case, a smoothing accelerated primal-dual mirror dynamical approach (SAPDMD) for the problem (1.1) is also proposed based on the smoothing approximation technique and APDMD, and it has accelerated convergence rates of Lagrangian and objective function gaps.
We provide a comparative explanation between SAPDMD and state-of-the-art dynamical approaches based on differential inclusion, Moreau-Yosida regularization and directional derivative methods. Moreover, we further analyze the convergence properties of SAPSMD and provide smooth parameter selection condition. Last, applying the SAPDMD to DCCP and EDMO leads to two distributed SAPDMDs (SADPDMD and SADMD) with accelerated convergence guarantees.

Last but not least, the asymptotic analysis and the obtained results in this paper can be straightforwardly transferred to inertial dynamical approaches for the problems (1.3), (1.5) and (1.7).

The paper is organized as follows. Section 2 introduces some preliminaries of convex analysis, saddle point theorem, dual distance, Bregman divergence, projection operators, graph theory and smoothing approximation. Then, we propose the APDMD to solve problem (1.1) in the smooth case, and discuss some convergence properties of APDMD in Section 3. In Section 4, we propose a SAPDMD (smoothing version of APDMD) for solving the problem (1.1) in the nonsmooth case, and provide a detailed discussion of the convergence properties of SAPDMD and then extend the SAPDMD to address the DCCP and EDMO in the nonsmooth case. Section 5 provides several experiments to demonstrate the theoretical results. Finally, we conclude this paper in Section 6.

2. Preliminaries. This section introduces some essential mathematical preliminaries.

2.1. Convex analysis. A function $f : X \to \mathbb{R}$ is convex, if it satisfies $\theta f(x) + (1 - \theta) f(z) \geq f(\theta x + (1 - \theta) z), \forall x, z \in X, x \neq z$, and $\theta \in [0, 1]$. The subdifferential $g_f(x)$ of $f(x)$ with respect to $x \in X$ is defined by $g_f(x) = \{ p \in \mathbb{R}^n | f(z) - f(x) \geq p^T(z - x), \forall x, z \in X \}$, and the element $\partial f(x)$ of $g_f(x)$ is called subgradient of $f(x)$.

In addition, if $f$ is smooth, the subgradient $\partial f(x)$ reduces to gradient $\nabla f(x)$.

2.2. Saddle point Theorem. The (augmented) Lagrangian $L_\beta : X \times \mathbb{R}^m \to \mathbb{R}$ with respect to the problem (1.1) with $\beta \geq 0$ is defined by

$$L_\beta(x, \lambda) = f(x) + \lambda^T(Ax - b) + \frac{\beta}{2} \|Ax - b\|^2,$$

Then, $(x^*, \lambda^*)$ is an optimal solution of the problem (1.1) if and only if it is a saddle point of $L_\beta$, i.e.,

$$L_\beta(x^*, \lambda) \leq L_\beta(x^*, \lambda^*) \leq L_\beta(x, \lambda^*), \forall (x, \lambda) \in X \times \mathbb{R}^n.$$

2.3. Dual distance and Bregman divergence. Define $\psi : X \to \mathbb{R}$ and let $X$ be a closed and convex set, then, by the Legendre-Fenchel transform, the conjugate function of $\psi$ is

$$\psi^*(u) = \sup_{x \in X} \{ u^T x - \psi(x) \}.$$

If $\psi(x)$ is a proper, lower semi-continuous and convex function, then for all $u \in X^*$, we have

$$\psi^{**}(u) = \sup_{u \in X^*} \{ x^T u - \psi^*(u) \} = \psi(x),$$

from the Fenchel’s duality theorem.
Assume that $\psi$ and $\psi^*$ are proper and convex functions, then they’re subdifferentiable, i.e., $\partial \psi(x)$, $\partial \psi^*(u)$ exist in the relative interior of their domains $\mathcal{X}$, $\mathcal{X}^*$, respectively [46]. In addition, using the condition that $\psi^{**}(u) = \psi(u)$, one has

$$\partial \psi^*(u) + \partial \psi(x) = u^T x \Leftarrow u \in \partial \psi(x) \Rightarrow x \in \partial \psi^*(u),$$

which implies $\partial \psi^*(u) = \arg\max_{x \in \mathcal{X}} \{ u^T x - \psi(x) \}$. Since $\text{dom} \psi = \mathcal{X}$, then $\partial \psi^*(u) \subset \mathcal{X}$, i.e., the set-valued mapping $\partial \psi^*$ can map $\mathcal{X}^*$ into $\mathcal{X}$. In order to make $\partial \psi^*(u) \subset \mathcal{X}$ be unique, i.e., it maps to a single-value, in other words, the $\psi^*(u)$ is differentiable for any $u \in \mathcal{X}^*$, the following definitions are needed [35].

**Definition 2.1.** A convex function $\psi$ is cofinite if its epigraph does not consist of any non-vertical half-line.

**Definition 2.2.** A convex function $\psi$ is essentially strictly convex if it is strictly convex and subdifferentiable on all convex subsets.

**Lemma 2.3.** If $\psi$ and $\psi^*$ are proper, convex, and closed, so they are inverses of each other, then $\psi^*$ is finite and differentiable on $\mathcal{X}^*$ if and only if $\psi$ is essentially strictly convex and cofinite.

**Lemma 2.4.** Bregman divergence: The distance between $h$ at $y \in \mathcal{X}$ and its first-order Taylor series approximation at $z \in \mathcal{X}$ is given by:

$$D_h(y, z) = h(y) - h(z) - \nabla h(z)^T (y - z), \forall y, z \in \mathcal{X},$$

which is nonnegative if $h$ is convex and is an approximation to the Hessian metric, i.e., $D_h(y, z) \approx \frac{1}{2} (y - z)^T \nabla^2 h(x)(y - x), \forall y, z \in \mathcal{X}$ when $y$ is close to $z$.

**2.4. Projection operators.** Define the projection operator of a closed and convex set $\mathcal{X}$ at $x$ be $P_{\mathcal{X}}(u) = \arg \min_{z \in \mathcal{X}} \| u - z \|$.

**Lemma 2.5.** 1): If $\mathcal{X}$ is a box set, i.e., $\mathcal{X} = \{ u \in \mathbb{R}^n \mid u_i \leq u_i \leq \bar{u}_i, \ i = 1, ..., n \}$, then, for any $i$, we have

$$P_{\mathcal{X}}(u_i) = \begin{cases} u_i, & u_i < u_i, \\ u_i, & u_i \leq u_i \leq \bar{u}_i, \\ \bar{u}_i, & u_i > \bar{u}_i. \end{cases}$$

2): If $\mathcal{X}$ is a sphere set, i.e., $\mathcal{X} = \{ u \in \mathbb{R}^n \mid \| u - z \| \leq r, z \in \mathbb{R}^n, r > 0 \}$, then, the projection operator is $P_{\mathcal{X}}(u) = \left\{ \begin{array}{ll} u, & \| u - z \| \leq r, \\ u + r(u - z), & \| u - z \| > r. \end{array} \right.$

3): If $\mathcal{X}$ is an affine set, i.e., $\mathcal{X} = \{ u \in \mathbb{R}^n \mid Au = b, A \in \mathbb{R}^{m \times n} \}$, then, the projection operator is $P_{\mathcal{X}}(u) = u + A^\dagger (b - Au)$, where $A^\dagger$ is Moore-Penrose pseudoinverse of $A$ if $\text{Rank}(A) < m$, and $A^\dagger = A^T (AA^T)^{-1}$ when $\text{Rank}(A) = m$.

4): If $\mathcal{X}$ is a half-space set, i.e., $\mathcal{X} = \{ u \in \mathbb{R}^n \mid a^T u \leq b, a \neq 0 \}$, then, the projection operator is $P_{\mathcal{X}}(u) = \left\{ \begin{array}{ll} u + \frac{b - a^T u}{\| a \|^2} a, & a^T u > b, \\ u, & a^T u \leq b. \end{array} \right.$

For more information regarding the projection operators, please refer to [10, 43].

**2.5. Graph Theory.** [26] An undirected communication topology graph is a triplet $G = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with node set $\mathcal{V} = \{ \nu_1, \nu_2, ..., \nu_n \}$, edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ and
connection matrix $A = \{a_{ij}\}_{i,j=1}^n$ with nonnegative elements $a_{ij} = a_{ji} > 0$ if $(i, j) \in \mathcal{E}$, and $a_{ii} = a_{ii} = 0$ otherwise. The coupling of agents in an undirected graph is unordered, which means that there exists information exchange for both agent $i$ and agent $j$. A path in an undirected graph between agent $i$ and agent $j$ is a sequence of edges of the form $(i, i_1), (i_1, i_2), \ldots, (i_s, j)$, where $i, i_1, \ldots, i_s, j$ denote different agents. Let $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$ be an agent $i$'s neighbor set. The undirected graph $G$ is connected if there exists a path between any pair of distinct nodes $v_i$ and $v_j$ ($i, j = 1, 2, \ldots, n$).

2.6. Smoothing approximation. The main characteristic of the smoothing method is to approximate the nonsmooth function with a parameterized smoothing function. In this paper, we adopt a smoothing function, which is defined as follows:

**Definition 2.6.** [14, 21] Let $\hat{f}: \mathcal{X} \times [0, \bar{\mu}] \rightarrow \mathbb{R}$ with $\bar{\mu} > 0$ be a smoothing function of the convex function $f: \mathcal{X} \rightarrow \mathbb{R}$, if $f(x, \mu)$ is continuous differentiable for any fixed $\bar{\mu} \geq \mu > 0$, then it enjoys the following properties:

(i) (approximation property) \(\lim_{x \rightarrow w, \mu \rightarrow 0} \hat{f}(x, \mu) = f(w), \forall x \in \mathcal{X}\),

(ii) (convexity) For any fixed $\mu > 0$, $\hat{f}(x, \mu)$ is a convex function of $x$ in $\mathcal{X}$,

(iii) (gradient consistency) \(\lim_{x \rightarrow w, \mu \rightarrow 0} \nabla_x \hat{f}(x, \mu) \subseteq \partial f(w), \forall x \in \mathcal{X}\),

(iv) (gradient boundedness and Lipschitz continuous of $\mu$) There exists a positive constant $\kappa_f > 0$ such that

\[
|\nabla_\mu \hat{f}(x, \mu)| \leq \kappa_f \quad \forall \mu \in [0, \bar{\mu}], \forall x \in \mathcal{X},
\]

and for any $x \in \mathcal{X}, u_1, u_2 \in [0, \bar{\mu}]$, it follows

\[
|\hat{f}(x, u_1) - \hat{f}(x, u_2)| \leq \kappa_f |u_1 - u_2|.
\]

(v) (Lipschitz continuity with respect to $x$) there is a constant $\ell$ such that for any fixed $\mu \in [0, \bar{\mu})$, $\nabla_x \hat{f}(x, \cdot)$ is Lipschitz continuous with respect to $x$ on $\mathcal{X}$ with a Lipschitz constant $\frac{\ell}{\mu}$.

In addition, the Definition 2.6 (iii) implies that $|\hat{f}(x, \mu) - f(x)| \leq \kappa_f \mu, \forall 0 < \mu \leq \bar{\mu}$, $x \in \mathcal{X}$.

The smoothing function satisfying the above conditions in Definition 2.6 enjoys the following properties:

**Lemma 2.7.** [13] 1): If $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n$ are smoothing functions of $f_1, f_2, \ldots, f_n$, then $\sum_{i=1}^n c_i \hat{f}_i$ is a smoothing function of $\sum_{i=1}^n c_i f_i$ with $\kappa_{\sum_{i=1}^n c_i f_i} = \sum_{i=1}^n c_i \kappa_{\hat{f}_i}$ when $c_i \geq 0$ and $f_i$ is regular for any $i = 1, 2, \ldots, n$.

2): If $\Phi: \mathcal{X} \rightarrow \mathbb{R}$ is locally Lipschitz, $\Psi: \mathbb{R} \rightarrow \mathbb{R}$ is continuously differentiable and globally Lipschitz with a Lipschitz constant $l_\Phi$. Let $\Phi$ be a smoothing function of $\Phi$, then $\Psi(\Phi)$ is a smoothing function of $\Phi$ with $\kappa_{\Psi(\Phi)} = l_{\Phi} \kappa_{\Phi}$.

3): Let $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}^n$ be regular and $\Psi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be continuously differentiable. If $\Phi$ is a smoothing function of $\Phi$, then $\Psi(\Phi)$ is a smoothing function of $\Phi(\Psi)$ with $\kappa_{\Psi(\Phi)} = \kappa_{\Phi(\Psi)}$.

**Example 2.8.** The existing results in [21] provide some theoretical basis to construct smoothing functions that satisfy the conditions in Definition 2.6. A smoothing
function for the $g(s) = \max \{0, s\}$ is given by

$$g(\mu, s) = \begin{cases} \max \{0, s\}, & \text{if } |s| > \mu, \\ \frac{(s + \mu)^2}{4\mu}, & \text{if } |s| \leq \mu, \end{cases}$$

with a $\kappa_{\hat{g}} = \frac{1}{4}$. Note that $\hat{g}(\mu, s)$ is convex and nondecreasing of $s$ with any fixed $0 < \mu \leq \hat{\mu}$, is also nondecreasing with respect to $\mu$ for any fixed $s$, and $\lim_{\mu \to 0} \hat{g}(\mu, s) = \max \{0, s\}$.

Many nonsmooth convex functions in applications can be reformulated by the max $\{s, 0\}$, such as

$$\max \{s, x\} = s + \max \{s - x, 0\},$$
$$\min \{s, x\} = s - \max \{s - x, 0\},$$
$$\max \{s, x, w\} = \min \{\max \{x, s\}, w\},$$
$$\theta(s) = |s| = \max \{s, 0\} + \max \{-s, 0\},$$

where the smoothing approximation function of $\theta(s) = |s|$ is

$$\hat{\theta}(s, \mu) = \begin{cases} |s|, & \text{if } |s| > \frac{\mu}{2}, \\ \frac{s^2}{\mu} + \frac{\mu}{2}, & \text{if } |s| \leq \frac{\mu}{2}, \end{cases}$$

where $\lim_{\mu \to 0} \hat{\theta}(s, \mu) = |s|$, and $\kappa_{\hat{\theta}} = \frac{1}{4}$ (It is used in Subsection 5.2).

2.7. Notation. Let $\mathbb{R}^n$ ($\mathbb{R}^{m \times m}$) be the $n$-dimensional (or $m$-by-$n$) real vectors (or real matrices), and $I_n$ be a $n \times n$ identity matrix. For vectors $x, y \in \mathbb{R}^n$, $x^Ty = \sum_{i=1}^{n} x_i y_i$, and the superscript $T$ represents transpose. col $(x_1, ..., x_n) = (x_1^T, ..., x_n^T)^T$, $\|\cdot\|$ is the Euclidean norm, $\|\cdot\|_1$ is the 1-norm. $\text{L}_{\text{loc}}^p ([t_0, +\infty))$ represents the local Lebesgue integrable functions on $[t_0, +\infty)$. Let $A_i \in \mathbb{R}^{p_i \times q_i}$ $(p_i, q_i > 0), i = 1, ..., n$, then we define $\tilde{A} = \text{bldiag} \{A_1, ..., A_n\} \in \mathbb{R}^{\sum_{i=1}^{n} p_i \times \sum_{i=1}^{n} q_i}$, which means a block diagonal matrix. For $x_i \in \mathbb{R}^{m_i}$, $i = 1, ..., n$, we use col $(x_1, ..., x_n) \in \mathbb{R}^{mn}$ to denote an $mn$ column vector.

3. Optimization approaches for problem (1.1) in the smooth case. In this section, we propose an APDMD approach to address the problem (1.1) with smooth and convex objective function. Then, we extend it to solve ECCP (3.1) and DEMO (3.2) in the smooth case, to obtain ADPDMD (3.20) and ADMD (3.25), respectively. To our best knowledge, there are no accelerated mirror dynamical approaches for the problem (1.1), distributed accelerated dynamical approaches for ECCP (3.1) and DEMO (3.2) only with smooth convex objective functions.

To make the well-posedness of the problem (1.1), some appropriate assumptions are needed, which are fairly standard.

Assumption 3.1.
1. The objective function $f(x)$ is smooth and convex on an open set containing $\mathcal{X}$, where $\mathcal{X}$ is a closed and convex set,
2. The function $\psi$ is proper, essentially strictly convex and cofinite (see Subsection 2.3),
3. (Slater’s condition) There exists a vector $x \in \text{int}(\mathcal{X})$ that satisfies $Ax = b$.

The formulation (1.1) covers two important network optimization problems as follows:
Scenario 1: Distributed Constrained Consensus Problem (DCCP). Consider a network of \( n \) agents over an undirected graph \( \mathcal{G} \). The aims of it is to cooperate for seeking the minimum of the following problem:

\[
\begin{align*}
\min f(x) &= \sum_{i=1}^{n} f_i(x_i), \\
\text{s.t. } Lx &= 0 \text{ (i.e., } x_i = x_j, i = 1, \ldots, n), x_i \in \mathcal{X}_i \subset \mathbb{R}^m,
\end{align*}
\]

where \( x = \text{col} (x_1, \ldots, x_n) \in \mathbb{R}^{nm} \), and \( \mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i \subset \mathbb{R}^{nm} \). \( L = L_n \otimes I_m \in \mathbb{R}^{n \times nm} \) and \( L_n \in \mathbb{R}^{n \times n} \) is the Laplacian matrix of \( \mathcal{G} \). \( Lx = 0 \) is applied to ensure the consensus of \( x_i = x_j, i, j = 1, 2, \ldots, n \) in a distributed way since the agent \( i \) only accesses its local function \( f_i : \mathcal{X}_i \to \mathbb{R} \) and the constraint \( \mathcal{X}_i \in \mathbb{R}^m \).

Scenario 2: Distributed Extended Monotropic Optimization (DEMO). Consider a network of \( n \) agents reciprocating information over a graph \( \mathcal{G} \). There exists a local objective function \( f_i(x_i) : \mathcal{X}_i \to \mathbb{R} \) and a local feasible constraint set \( \mathcal{X}_i \subset \mathbb{R}^{p_i}, i = 1, \ldots, n \). Let \( x = \text{col} (x_1, \ldots, x_n) \), \( \mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i \subset \mathbb{R}^{\sum_{i=1}^{n} p_i} \), then, the DEMO problem is

\[
\begin{align*}
\min f(x) &= \sum_{i=1}^{n} f_i(x_i), \\
\text{s.t. } \sum_{i=1}^{n} A_i x_i &= \sum_{i=1}^{n} d_i, x_i \in \mathcal{X}_i, i = 1, 2, \ldots, n,
\end{align*}
\]

where \( A = [A_1, \ldots, A_n] \in \mathbb{R}^{m \times \sum_{i=1}^{n} p_i} \) and \( A_i \in \mathbb{R}^{m \times p_i} \).

To ensure the well-posedness of DCCP (3.1) and DEMO (3.2), some appropriate assumptions need to be made on it, which are fairly standard.

**Assumption 3.2.**
1. The objective function \( f(x) = \sum_{i=1}^{n} f_i(x_i) \) and for all \( i = 1, \ldots, n \), \( f_i(x_i) \) is smooth and convex on an open set containing \( \mathcal{X}_i \), and \( \mathcal{X}_i \) is a closed and convex set,
2. For all \( i = 1, \ldots, n \), \( \psi_i \) is proper, essentially strictly convex and cofinite (see Subsection 2.3),
3. The communication graph \( \mathcal{G} \) is connected and undirected,
4. The Slater’s condition of DCCP (3.1) and DEMO (3.2) is satisfied.

3.1. APDMD for problem (1.1) with smooth convex objective functions. Inspired by the accelerated mirror descent [36] and primal-dual dynamical approach [25], we propose the following APDMD:

\[
\begin{align*}
\dot{x}(t) &= \frac{\alpha}{t} (\nabla \psi^* (u(t)) - x(t)), \\
\dot{u}(t) &= -\frac{\alpha}{t} (\nabla f(x(t)) + \beta A^T (Ax(t) - b) + A^T v(t)), \\
\dot{\lambda}(t) &= \frac{\alpha}{t} (v(t) - \lambda(t)), \\
\dot{v}(t) &= \frac{\alpha}{t} (A \nabla \psi^* (u(t)) - b), \\
x(t_0) &= x_0, u(t_0) = u_0 \text{ with } \nabla \psi^* (u_0) = x_0 \in \mathcal{X}, \lambda(t_0) = \lambda_0, v(t_0) = v_0,
\end{align*}
\]

where \( t \geq t_0 > 0 \), \( \beta > 0 \), and \( \alpha \geq 2 \). The illustration of the APDMD (3.3) is in Figure 1 (left).

Note that the APDMD (3.3) can be equivalently transformed into the following second-order dynamical approach:
\[
\begin{align*}
\dot{x}(t) + \frac{\alpha + 1}{t} \dot{x}(t) + \nabla^2 \psi^* \left( \nabla \psi \left( x(t) + \frac{t}{\alpha} \dot{x}(t) \right) \right) \\
	imes \left( \nabla f(x(t)) + \beta A^T (Ax(t) - b) + A^T \left( \lambda(t) + \frac{t}{\alpha} \dot{\lambda}(t) \right) \right) &= 0, \\
\ddot{\lambda}(t) + \frac{\alpha + 1}{t} \dot{\lambda}(t) - A \left( x(t) + \frac{t}{\alpha} \dot{x}(t) \right) + b &= 0, \\
x(t_0) &= x_0 \in \mathcal{X}, \dot{x}(t_0) = \dot{x}_0, \lambda(t_0) = \lambda_0, \dot{\lambda}(t_0) = \dot{\lambda}_0,
\end{align*}
\]

where the Hessian term \( \nabla^2 \psi^* \left( \nabla \psi \left( x(t) + \frac{t}{\alpha} \dot{x}(t) \right) \right) \) is nonlinear transformation. It applies to \( \nabla f(x(t)) + \beta A^T (Ax(t) - b) + A^T \left( \lambda(t) + \frac{t}{\alpha} \dot{\lambda}(t) \right) \) to guarantee the trajectory of \( x \) is inside the feasible set \( \mathcal{X} \) in the intuitive understanding (the rigorous proof process in Lemma 3.3). The form in (3.4) allows us to understand more intuitively why the proposed (3.3) is called the primal-dual method that is with primal variable \( x \) and dual variable \( \lambda \).

**3.2. Interpretation of the APDMD.** The APDMD (3.3) (i.e., (3.4)) can be interpreted from different perspectives: neurodynamic approach, Hamilton’s system, game theory and control theory.

- **Neurodynamic approach perspective.** The APDMD (3.3) be regarded as a neurodynamic approach (describe the dynamic behavior of neurons), thus APDMD can be implemented by analog circuit like the Hopfield neural network [32], as shown in Figure 2. The circuit in Figure 2 is designed according to APDMD (3.3) by using analog adders, analog subtracters, analog multipliers, analog integrators, etc. When the circuit is turned on, the stable value of the voltage at position \( x \) in Figure 2 is the stable point to the APDMD (3.3) (i.e., an optimal solution to the problem (1.1)). For more information about neurodynamic approaches and their circuits, please see [33, 52].

- **Hamiltonian system perspective.** The Hamiltonian-based system approach is used to design accelerated dynamical approaches for solving problem (1.3), which has recently been widely studied in [22, 48], and it plays a key role in designing of APDMD (3.3) and the development of our Lyapunov analysis method.
Designing the first Hamiltonian (time-dependent) system

\[
H_1 (\ddot{x}(t), u(t), \eta(t)) = \psi^*(u(t)) + \Gamma (\eta(t)) \left( f \left( \frac{\ddot{x}(t)}{\eta(t)} \right) \right) + \beta \left\| A \left( \frac{\ddot{x}(t)}{\eta(t)} \right) - b \right\|^2 + v(t)^T \left( A \left( \frac{\ddot{x}(t)}{\eta(t)} \right) - b \right)
\]

with \( \ddot{x}(t) = \eta(t) x(t) \), \( \eta(t) = t^\alpha \) and \( \Gamma (\eta(t)) = \frac{(\eta(t))^2}{2} \).

The corresponding continuous-time dynamics of \( H_1 \) is

\[
\frac{d}{dt} \left( \ddot{x}(t) \right) = \ddot{\eta}(t) \frac{d\ddot{x}(t)}{d\eta(t)} = \ddot{\eta}(t) \nabla_u H_1 (\ddot{x}(t), u(t), \eta(t)) = \ddot{\eta}(t) \nabla \psi^* (u(t)) \Rightarrow \dot{x}(t) = \frac{\alpha}{t} (\nabla \psi^* (u(t)) - x(t));
\]
\[
\frac{d}{dt} u(t) = \dot{\eta}(t) \frac{du(t)}{d\eta(t)} = -\Gamma(\eta(t)) \dot{\eta}(t) \nabla_x H_1(\bar{x}(t), u(t), \eta(t))
\]
\[
\Rightarrow \dot{u}(t) = -\frac{\alpha}{t} \left( \nabla f(x(t)) + \beta A^T(Ax(x(t)) - b) + A^T v(t) \right).
\]

Furthermore, setting the second Hamiltonian (time-dependent) system be
\[
H_2(\bar{\lambda}(t), v(t), \eta(t)) = \frac{1}{2} \|v(t)\|^2 - \Gamma(\eta(t)) \left( \frac{\bar{\lambda}(t)}{\eta(t)} \right)^T (Au(t) - b),
\]
where \( \bar{\lambda}(t) = \eta(t) \lambda(t) \).

The corresponding continuous-time dynamics of \( H_2 \) is
\[
\frac{d}{dt} (\bar{\lambda}(t)) = \dot{\eta}(t) \frac{d\bar{\lambda}(t)}{d\eta(t)} = \dot{\eta}(t) \nabla_\bar{\lambda} H_2(\bar{\lambda}(t), v(t), \eta(t)) = \dot{\eta}(t) v(t)
\]
\[
\Rightarrow \dot{\bar{\lambda}}(t) = \frac{\alpha}{t} (v(t) - \lambda(t))
\]
\[
\frac{d}{dt} v(t) = \dot{\eta}(t) \frac{dv(t)}{d\eta(t)} = -\Gamma(\eta(t)) \dot{\eta}(t) \nabla_\lambda H_2(\bar{\lambda}(t), v(t), \eta(t))
\]
\[
\Rightarrow \dot{v}(t) = -\frac{\alpha}{\bar{\lambda}(t)} (b - Au(t)).
\]

From the above conclusions, the APDMD (3.3) can be obtained directly.

- **Game theoretic standpoint** [4]. Let us consider \( x(t) \) and \( \lambda(t) \) as two players who compete with each other. Briefly, we identify players by their actions. We can see that each player anticipates the opponent’s movement in APDMD (3.4). In the coupling term, the player \( \lambda(t) \) takes account of the anticipated position of the player \( x(t) \), i.e., \( x(t) + \frac{\alpha}{t} \dot{x}(t) \). Conversely, for player \( x(t) \), it takes account of the anticipated position of the player \( \lambda(t) \), i.e., \( \lambda(t) + \frac{\lambda}{t} \dot{\lambda}(t) \).

- **Control theoretic view.** The APDMD (3.4) can also be associated with control theory and state derivative feedback. Let \( \chi(t) = \text{col}(x(t), \lambda(t)) \), the APDMD (3.4) can be written in the following formulation
\[
\ddot{\chi}(t) + \frac{\alpha + 1}{t} \dot{\chi}(t) = \Upsilon(t, \chi(t), \dot{\chi}(t)),
\]
with an operator \( \Upsilon \), i.e., a feedback control term which takes the constraint into account. It is a function of the state \( \chi(t) \), its derivative \( \dot{\chi}(t) \) and time \( t \).

For a comprehensive understanding of state derivative feedback, the readers can consult reference [40].

### 3.3. Feasibility, existence and uniqueness of strong global solution to APDMD.

In this subsection, we illustrate the feasibility, the existence and uniqueness of the strong global solution \( x(t) \) for APDMD (3.3) by the Cauchy-Lipschitz-Picard theorem [15].

**Lemma 3.3.** For any initial values \( (x(t_0), u(t_0), \lambda(t_0), v(t_0)) \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \), the variable \( x(t) \in \mathcal{X} \), \( \forall t \geq t_0 > 0 \), i.e., the solution \( x(t) \) is feasible.

**Proof.** Inspired by the work in [35], we provide a rigorous proof for the feasibility by contradiction. Suppose there exist \( t_1 > 0 \) and \( x(t_1) \notin \mathcal{X} \). Since \( \mathcal{X} \) is closed and convex, by the separation theorem, there exists a hyperplane that strictly separates \( x(t_1) \) and the set \( \mathcal{X} \). That is, there exist \( \omega, a \in \mathbb{R}^n \) such that

\[
\omega \cdot x(t_1) + a > 0.
\]
\((x(t_1) - a)^T \omega > 0\) and \((x-a)^T \omega < 0\), \(\forall x \in \mathcal{X}\). Let \(d(x(t)) = (x-a)^T \omega\). Note that the trajectory \(x(t)\) is continuous, then \(t \rightarrow d(x(t))\) is continuous, and \(d(x(t_0)) = (x(t)-a)^T \omega < 0\), \(d(x(t_1)) = (x(t_1) - a)^T \omega > 0\), thus, there is a \(t_2\) such that \(d(x(t_2)) = 0\), i.e., \(d(x(t)) > 0\), \(t \in (t_2, t_1)\) and \(d(x(t)) < 0\), \(t \in [t_0, t_2)\), which implies \(t_2 = \text{sub} \{ t : d(x(t)) \leq 0 \}\). We have \(d(x(t_1)) - d(x(t_0)) > 0\), according to Taylor’s theorem and \(\dot{x}(t) = \frac{\alpha}{\tau} (\nabla \psi^*(u(t)) - x(t))\), there exists \(\tau \in [t_2, t_1]\), such that

\[
d(x(t_1)) - d(x(t_2)) = \dot{d}(x(t)) \frac{\alpha}{\tau} (\nabla \psi^*(u(t)) - x(t)) < 0,
\]

since \(d(\nabla \psi^*(u_0)) \in \mathcal{X}\). It leads to a contradiction, which concludes the proof and is shown in Figure 1 (right).

**Definition 3.4.** Let \(x : [t_0, +\infty) \rightarrow \mathcal{X}, \lambda : [t_0, +\infty) \rightarrow \mathbb{R}^m\) with the corresponding product space structure \(\mathcal{X} \times \mathbb{R}^m\) and \(t_0 > 0\). \((x, u, \lambda, v)\) is a strong global solution of APDMD (3.3) if it satisfies the following properties:

(i) \(x, u : [t_0, +\infty) \rightarrow \mathbb{R}^n, \lambda, v : [t_0, +\infty) \rightarrow \mathbb{R}^m\) are locally absolutely continuous,

(ii) The APDMD holds for almost every \(t \geq t_0\),

(iii) \(x(t_0) = x_0 \in \mathcal{X}, u(t_0) = u_0, \lambda(t_0) = \lambda_0, v(t_0) = v_0\).

**Theorem 3.5.** For any initial values \((x(t_0), u(t_0), \lambda(t_0), v(t_0)) \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m\), there exists a unique strong global solution of the APDMD.

**Proof.** Let \(Y(t) = (x(t), u(t), \lambda(t), v(t))\), then the APDMD can be rewritten as follows:

\[
\begin{align*}
\dot{Y}(t) &= F(t, Y(t)) \\
Y(t_0) &= (x(t_0), u(t_0), \lambda(t_0), v(t_0)),
\end{align*}
\]

where \(F : [t_0, +\infty) \times \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m, F(t, Y) = (\frac{\alpha}{\tau} (\nabla \psi^*(u) - x), -\frac{\alpha}{\tau} (\nabla f(x) + \beta A^T (Ax - b) - A^T v), \frac{\alpha}{\tau} (v - \lambda), \frac{\alpha}{\tau} (A \nabla \psi^*(u) - b), \frac{\alpha}{\tau} (v - \lambda), \frac{\alpha}{\tau} (A \nabla \psi^*(u) - b)).\)

According to the Cauchy-Lipschitz-Picard theorem, there exists a unique strong global solution for APDMD (3.3), i.e., (3.4), if the following conditions (I) and (II) are fulfilled. (I): For every \(t \in [t_0, +\infty)\), the mapping \(F(t, \cdot)\) is \(l(t)\)-Lipschitz continuous and \(l(\cdot) \in L^1_{loc}(t_0, +\infty)\).

(II): For any \(Y \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m\), we have \(F(\cdot, Y) \in L^1_{loc}([t_0, +\infty), \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m)\).

The proof of (I). Let \(t \in [t_0, +\infty)\) be fixed and utilize the Lipschitz continuous of \(\nabla \psi^*, \nabla f\) and \(\|X + Y\|^2 \leq 2\|X\|^2 + 2\|Y\|^2\) for any vectors \(X\) and \(Y\), then, for any \(Y\) and \(\dot{Y}\), we have

\[
\left\| F(t, Y) - F(t, \dot{Y}) \right\| \
\leq \left( \frac{t^2}{\alpha^2} \left( 4\beta^2 \delta_{\max}(A^T A)^2 + 4t_f + \delta_{\max}(A^T A) t_v^* \right) \right)^{\frac{1}{2}} \left\| Y - \dot{Y} \right\|,
\]

where \(\delta_{\max}(A)\) is the maximum singular value of matrix \(A\) and the inequality holds from the Lipschitz continuous properties of \(\nabla f\) and \(\nabla \psi^*\) that have Lipschitz constants.
Theorem 3.4. Convergence rate of APDMD. With the help of the Lyapunov analysis method with Bregman divergence function, we will illustrate the accelerated convergence properties of the APDMD (3.3) in the following.

Theorem 3.6. Suppose Assumption 3.1 holds. Let $(x(t), \lambda(t))$ and $(x^*, \lambda^*)$ be the solution trajectory and optimal solution for APDMD (3.4) and problem (1.1), respectively. Then for any $(x(t_0), u(t_0), \lambda(t_0), v(t_0)) \in \mathcal{X} \times \mathbb{R}^m \times \mathbb{R}^m$ with $\alpha \geq 2$, the following statements are true:

(I): The trajectories $x(t)$ and $\lambda(t)$ of APDMD (3.4) are bounded for any $t \geq t_0 > 0$.

(II): Let $\mathcal{L}_\beta (x(t), \lambda(t)) = f(x(t)) + \lambda(t)^T(Ax(t) - b) + \frac{\beta}{2} \|Ax(t) - b\|^2$ with $\beta \geq 0$. Then, we have

\begin{align*}
\mathcal{L}_\beta (x(t), \lambda(t)) - \mathcal{L}_\beta (x^*, \lambda(t)) &\leq \frac{\alpha^2 V(t_0)}{t^2}, \\
\|Ax(t) - b\|^2 &\leq \frac{2\alpha^2 V(t_0)}{\beta t^2}, \\
\int_0^{+\infty} t (\mathcal{L}_\beta (x(t), \lambda^*) - \mathcal{L}_\beta (x^*, \lambda(t))) \, dt &< +\infty, \\
\int_{t_0}^{+\infty} t \|Ax(t) - b\|^2 \, dt &< +\infty.
\end{align*}
(III): Let $\lambda^* = \begin{cases} 0, & \text{if } Ax(t) - b = 0, \\ \frac{Ax(t) - b}{\|Ax(t) - b\|}, & \text{if } Ax(t) - b \neq 0, \end{cases}$. For any $t \geq t_0 > 0$, it follows that

\begin{align}
0 & \leq f(x(t)) - f(x^*) + \|Ax(t) - b\| + \frac{\beta}{2} \|Ax(t) - b\|^2 \leq \frac{\alpha^2 V(t_0)}{t^2}, \\
-\frac{\alpha \sqrt{V(t_0)}}{t} & \leq f(x(t)) - f(x^*) \leq \frac{\alpha^2 V(t_0)}{t^2}.
\end{align}

Proof. Consider a Lyapunov function $V : [t_0, T) \to [0, +\infty)$ which is given by

\begin{equation}
V(t) = V_1(t) + V_2(t) + V_3(t),
\end{equation}

where

\begin{align}
V_1(t) &= \frac{t^2}{2} (L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t))) \\
 &= \frac{t^2}{2} \left( f(x(t)) - f(x^*) + \frac{\beta}{2} \|Ax(t) - b\|^2 + A^T \lambda^* (x(t) - x^*) \right), \\
V_2(t) &= D_{\psi^*} (u(t), u^*) = \psi^* (u(t)) - \psi^* (u^*) - \nabla \psi^* (u^*)^T (u(t) - u^*), \\
V_3(t) &= D_h (v(t), \lambda^*) = \frac{1}{2} \|v(t) - \lambda^*\|^2,
\end{align}

the $D_{\psi^*} (u, u^*)$, $D_h (v, \lambda^*)$ are two Bregman divergences associated with suitable functions $\psi^*$ and $h$ that are determined by different requirements, respectively.

Note that the Lyapunov function $V_1(t)$ is positive and radially unbounded since $L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t)) \geq 0, \forall (x(t), \lambda(t)) \in X \times \mathbb{R}^m$ in (2.2), $\alpha \geq 2$ and $t \geq t_0 > 0$. For $V_2(t)$ and $V_3(t)$, since $\psi^*(\cdot)$ and $h(\cdot) = \frac{1}{2} \|\cdot\|^2$ are convex (Assumption 3.1 holds), we get $D_{\psi^*} (u(t), u^*) \geq 0, D_h (v(t), \lambda^*) \geq 0$, moreover, $\|u(t)\| \to +\infty, V_2(t) \to +\infty$ and $\|v(t)\| \to +\infty, V_3(t) \to +\infty$. Thus, both $V_2(t)$ and $V_3(t)$ are positive and radially unbounded.

The derivatives of $V_1(t)$, $V_2(t)$ and $V_3(t)$ along the trajectory of APDMD with $x(t) \in X$ are given by

\begin{align}
\dot{V}_1(t) &= \frac{2t}{\alpha^2} \left( L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t)) \right) + (\nabla \psi^* (u(t)) - x(t))^T \\
&\quad \times \frac{t}{\alpha} \left( \nabla f(x(t)) + \frac{\beta}{2} A^T (Ax(t) - b) + A^T \lambda^* \right) \\
&= \frac{2t}{\alpha^2} \left( f(x(t)) - f(x^*) + \frac{\beta}{2} \|Ax(t) - b\|^2 + (A^T \lambda^*)^T (x(t) - x^*) \right) \\
&\quad + \frac{t}{\alpha} \left( \nabla f(x(t)) + \beta A^T A (x(t) - x^*) + A^T \lambda^* \right)^T \\
&\quad \times (\nabla \psi^* (u(t)) - \nabla \psi^* (u^*)) + (\nabla \psi^* (u^*) - x(t))^T \\
&\quad + \frac{t}{\alpha} \left( \nabla f(x(t)) + \beta A^T A (x(t) - x^*) + A^T \lambda^* \right), \\
\dot{V}_2(t) &= (\nabla \psi^* (u(t)) - \nabla \psi^* (u^*))^T \dot{u}(t) \\
&= -\frac{t}{\alpha} \left( \nabla \psi^* (u(t)) - \nabla \psi^* (u^*) \right)^T \\
&\quad \times (\nabla f(x(t)) + \beta A^T (Ax(t) - b) + A^T v(t)) \\
&= -\frac{t}{\alpha} \left( \nabla \psi^* (u(t)) - \nabla \psi^* (u^*) \right)^T \\
&\quad \times (\nabla f(x(t)) + \beta A^T A (x(t) - x^*) + A^T \lambda^*) \\
&\quad - \frac{t}{\alpha} (\nabla \psi^* (u(t)) - \nabla \psi^* (u^*))^T (A^T v(t) - A^T \lambda^*),
\end{align}
\[ \dot{V}_3(t) = (v(t) - \lambda^*)^T \dot{v} = \frac{t}{\alpha} (v(t) - \lambda^*)^T A (\nabla \psi^* (u(t)) - x^*). \]

Adding (3.9), (3.10) and (3.11) together, and using \( Ax^* = b \), we have
\[
\dot{V}(t) = \dot{V}_1(t) + \dot{V}_2(t) + \dot{V}_3(t)
\]
\[
= \frac{2t}{\alpha^2} \left( f(x(t)) - f(x^*) + \frac{\beta}{2} \| Ax(t) - b \|^2 + \left( A^T \lambda^* \right)^T (x(t) - x^*) \right)
+ \frac{t}{\alpha} \left( \nabla f(x(t)) + \frac{\beta}{2} A^T (Ax(t) - b) + A^T \lambda^* \right)^T (x^* - x(t))
\leq \frac{2t}{\alpha^2} \left( L_\beta(x(t), \lambda^*) - L_\beta(x^*, \lambda^*) \right) - \frac{t}{2\alpha} \left( L_\beta(x^*, \lambda^*) - L_\beta(x(t), \lambda(t)) \right)
- \frac{\beta t}{2\alpha} (Ax(t) - b) A(x^* - x(t))
= - \frac{t}{\alpha} \left( L_\beta(x(t), \lambda^*) - L_\beta(x^*, \lambda(t)) \right) - \frac{\beta t}{2\alpha} \| Ax(t) - b \|^2 \leq 0,
\]

where the second equation holds from \( \nabla \psi^*(u^*) = x^* \), the first inequality is satisfied due to the convexity of \( L_\beta(x(t), \lambda^*) \) of \( x(t) \) with fixed \( \lambda^* \), the third equation holds because of \( L_\beta(x^*, \lambda(t)) = L_\beta(x^*, \lambda^*) \) and the last inequality is established since \( \alpha \geq 2 \) and \( L_\beta(x(t), \lambda^*) \) is bounded for any \( t \geq t_0 > 0 \).

Since the Lyapunov function \( V(t) \) is positive and radially unbounded with any \( t \geq t_0 > 0 \), this implies that the trajectories of \( x(t) \) and \( \lambda(t) \) are bounded for any \( t \geq t_0 > 0 \). The proof of (I) therefore completed.

(II): From (3.12), we have that \( \dot{V}(t) \leq 0 \), i.e., \( V(t) \) is nonincreasing on \([t_0, +\infty) \). Thus, for any \( t \geq t_0 > 0 \) it yields to \( \frac{\alpha - 2}{\alpha^2} \left( L_\beta(x(t), \lambda^*) - L_\beta(x^*, \lambda(t)) \right) = V_1(t) \leq V(t) \leq V(t_0) \), i.e., \( L_\beta(x(t), \lambda^*) - L_\beta(x^*, \lambda(t)) \leq \frac{\alpha - 2}{\alpha^2} \frac{\| Ax(t) - b \|^2}{\sqrt{t}} \), thus the condition (3.6a) holds.

In addition, the (3.6a) further implies \( \| Ax(t) - b \|^2 \leq \frac{2\alpha - 2}{\alpha \sqrt{t}} \), i.e., (3.6b) is satisfied.

Since \( \dot{V}(t) \leq -t \frac{\alpha - 2}{\alpha^2} \left( L_\beta(x(t), \lambda^*) - L_\beta(x^*, \lambda(t)) \right) - \frac{\beta t}{2\alpha} \| Ax(t) - b \|^2 \), thus, integrating the aboved inequality from \( t_0 \) to \(+\infty\) and rearranging it, we have
\[
\frac{\alpha - 2}{\alpha^2} \int_{t_0}^{+\infty} t \left( L_\beta(x(t), \lambda^*) - L_\beta(x^*, \lambda(t)) \right) dt + \frac{\beta t}{2\alpha} \int_{t_0}^{+\infty} t \| Ax(t) - b \|^2 dt
= V(t_0) - V(+\infty) \leq V(t_0) < +\infty,
\]
and it combines with \( L_\beta(x(t), \lambda^*) - L_\beta(x^*, \lambda(t)) \geq 0 \), \( \forall x(t) \in X \) leads to
\( \int_{t_0}^{+\infty} t \| Ax(t) - b \|^2 dt < +\infty \), i.e., the condition (3.6c) holds. Moreover, (3.6c) combines with \( \| Ax(t) - b \|^2 \geq 0 \) implies \( \int_{t_0}^{+\infty} t \| Ax(t) - b \|^2 dt < +\infty \), i.e., the (3.6d) holds.

(III): The equation \( \lambda^* = \begin{cases} 0, & \text{if } Ax(t) - b = 0, \\ \frac{Ax(t) - b}{\| Ax(t) - b \|}, & \text{if } Ax(t) - b \neq 0, \end{cases} \) and (3.6a) yields 0 \leq f(x(t)) - f(x^*) + \| Ax(t) - b \| + \frac{\beta t}{2} \| Ax(t) - b \|^2 \leq \frac{\alpha^2 V(t)}{\alpha - 2} \).

Observing that the upper-bound of (2.2) in Subsection 2.2 implies
\[
f(x(t)) - f(x^*) \geq - (Ax(t) - b)^T \lambda^* \]
\[
\geq - \| Ax(t) - b \| \geq - \frac{\alpha \sqrt{V(t_0)}}{t}.
\]
Combining the (3.13), \(|Ax(t) - b| \geq 0\) and \(f(x(t)) - f(x^*) \leq \frac{\alpha^2 Y(t_0)}{t_0^2}\), we deduce that \(-\frac{\alpha\sqrt{Y(t_0)}}{t_0} \leq f(x(t)) - f(x^*) \leq \frac{\alpha^2 Y(t_0)}{t_0}\) with any \(t \geq t_0 > 0\). Finally, the conclusion (III) holds.

3.5. Examples of the APDMD. In this subsection, we will illustrate some examples of the APDMD (3.3), i.e., (3.4), in other words, the APDMD can reduce to some classical and new dynamical approaches when choosing different constraint set \(X\).

Case 1: If \(X\) is an Euclidean space, i.e., \(X = \mathbb{R}^n\). Let \(\psi(x(t) + \frac{\alpha}{t} \dot{x}(t)) = \frac{1}{2} \|x(t) + \frac{u}{t} \dot{x}(t)\|^2\), then, one has \(\nabla \psi(x(t) + \frac{\alpha}{t} \dot{x}(t)) = u(t)\), \(\psi^*(u(t)) = \frac{1}{2} \|u(t)\|^2\), \(\nabla \psi^*(u(t)) = u(t)\), \(\nabla^2 \psi^*(\nabla \psi(x(t) + \frac{\alpha}{t} \dot{x}(t))) = I_n\). The APDMD (3.4) reduces to the classical accelerated primal-dual dynamical approaches [30, 50, 18, 4].

\[
\begin{align*}
\dot{x}(t) + \frac{\alpha + 1}{t} \dot{x}(t) + \frac{\alpha}{t} \dot{\lambda}(t) + A(x(t) + \frac{\alpha}{t} \dot{x}(t)) + b = 0, \\
\dot{\lambda}(t) + \frac{\alpha + 1}{t} \dot{\lambda}(t) - A(x(t) + \frac{\alpha}{t} \dot{x}(t)) + b = 0, \\
x(t_0) = x_0 \in X, \quad \dot{x}(t_0) = \dot{x}_0, \quad \dot{\lambda}(t_0) = \dot{\lambda}_0.
\end{align*}
\]

(3.14)

Case 2: [9] If \(X\) is a positive-orthant constrained set, i.e., \(X = \mathbb{R}^n_+\). Let the negative entropy function \(\psi(x) = \sum_{i=1}^n x_i \ln x_i\) be a distance generating function in this case. Then, one has \(\nabla \psi(x) = \text{col}(1 + \ln x_1, ..., 1 + \ln x_n)\) and \(\psi^*(u) = \sum_{i=1}^n e^{u_i-1}, \nabla \psi^*(u) = \text{col}(e^{u_1-1}, ..., e^{u_n-1})\). \(\nabla^2 \psi^*(u(t)) = \text{diag}(e^{u_1-1}, ..., e^{u_n-1})\). Thus, for all \(x \in \mathbb{R}^n_+\), we have

\[
\nabla^2 \psi^*(\nabla \psi\left(x(t) + \frac{t}{\alpha} \dot{x}(t)\right)) = \text{diag}\left(x_1(t) + \frac{t}{\alpha} x_1(t), ..., x_n(t) + \frac{t}{\alpha} x_n(t)\right) = \text{diag}\left(x(t) + \frac{t}{\alpha} \dot{x}(t)\right).
\]

Correspondingly, APDMD (3.4) is turned into

\[
\begin{align*}
\dot{x}(t) + \frac{\alpha + 1}{t} \dot{x}(t) + \frac{t}{\alpha} \dot{x}(t) \left(\nabla f(x(t)) + \beta A^T(Ax(t) - b) + A^T(\dot{\lambda}(t) + \frac{\alpha}{\lambda} \dot{\lambda}(t))\right) = 0, \\
\dot{\lambda}(t) + \frac{\alpha + 1}{t} \dot{\lambda}(t) - A(x(t) + \frac{\alpha}{t} \dot{x}(t)) + b = 0, \\
x(t_0) = x_0 \in X, \quad \dot{x}(t_0) = \dot{x}_0, \quad \dot{\lambda}(t_0) = \dot{\lambda}_0, \quad \dot{\lambda}_0 = \dot{\lambda}_0.
\end{align*}
\]

(3.15)

In addition, for the constrained set \(X = \mathbb{R}^n_+\), the distance function can also be selected as \(\psi(x) = -\sum_{i=1}^n x_i \ln x_i\). Then, we have \(\nabla \psi(x) = \text{col}\left(-\frac{1}{x_1}, ..., -\frac{1}{x_n}\right)\) and the Bregman divergence of \(\psi\) for \(x\) and \(y\) is \(D_{\psi}(x,y) = \sum_{i=1}^n \left(\frac{x_i}{y_i} \ln \frac{x_i}{y_i} - 1\right)\) (named as the Itakura-Saito divergence). Further we have \(\psi^*(u) = -\sum_{i=1}^n (1 + \ln (-u_i))\) on \(\mathbb{R}^n_+\), \(\nabla \psi^*(u) = \text{col}\left(-\frac{1}{u_1}, ..., -\frac{1}{u_n}\right)\) and \(\nabla^2 \psi^* = \text{diag}\left(\frac{1}{(u_1)^2}, ..., \frac{1}{(u_n)^2}\right)\) on \(\mathbb{R}^n_+\). To sum up, one has

\[
\nabla^2 \psi^*(\nabla \psi\left(x(t) + \frac{t}{\alpha} \dot{x}(t)\right)) = \text{diag}\left(x_1(t) + \frac{t}{\alpha} \dot{x}_1(t), ..., x_n(t) + \frac{t}{\alpha} \dot{x}_n(t)\right)^2.
\]

(3.16)
ACCELERATED PRIMAL-DUAL MIRROR

Case 3: [11] If \( \mathcal{X} \) is a unit simplex set, i.e., \( \mathcal{X} = \Delta = \{ x \in \mathbb{R}^n_+ | \sum_{i=1}^n x_i = 1 \} \). Considering a distance-generating function \( \psi (x) = \sum_{i=1}^n x_i \ln x_i + I_X (x) \), where \( \sum_{i=1}^n x_i \ln x_i \) is the negative entropy function and \( I_X (x) \) is the indicator function on \( \mathcal{X} \) and its Bregman divergence is \( D_\psi (x, y) = \sum_{i=1}^n x_i \ln \frac{x_i}{y_i} \) between the vectors \( x, y \) (named as Kullback-Leibler divergence). Then, we have

\[

\nabla \psi (x) = \text{col} (1 + \ln x_1, \ldots, 1 + \ln x_n), \forall x \in \mathcal{X}, \\
\psi^* (u) = \ln \left( \sum_{i=1}^n e^{u_i} \right), \nabla \psi^* (u) = \text{col} \left( \frac{e^{u_1}}{\sum_{i=1}^n e^{u_i}}, \ldots, \frac{e^{u_n}}{\sum_{i=1}^n e^{u_i}} \right), \\

\left[ \nabla^2 \psi^* (u) \right]_{ij} = \frac{e^{u_i} \sum_{i=1}^n e^{u_i} - e^{u_i} e^{u_j}}{\left( \sum_{i=1}^n e^{u_i} \right)^2}, \forall i, j \in n.

\]

Furthermore, for any \( i, j = 1, \ldots, n \), we have

\[

\left[ \nabla^2 \psi^* \left( \nabla \psi \left( x (t) + \frac{t}{\alpha} \hat{x} (t) \right) \right) \right]_{ij} = \left( x_i (t) + \frac{t}{\alpha} \hat{x}_i (t) \right) - \left( x_i (t) + \frac{t}{\alpha} \hat{x}_i (t) \right) \left( x_j (t) + \frac{t}{\alpha} \hat{x}_j (t) \right).

\]

Therefore, the APDMD (3.4) reduces to

\[

\begin{cases}
\dot{x}_i (t) = \frac{\alpha - 1}{\alpha} \hat{x}_i (t) + \left( x_i (t) + \frac{t}{\alpha} \hat{x}_i (t) \right) \left( \nabla f (x (t)) + \beta A^T (Ax (t) - b) + A^T \left( \lambda (t) + \frac{t}{\alpha} \hat{\lambda} (t) \right) \right) \\
\dot{\lambda} (t) = \frac{\alpha + 1}{\alpha} \hat{\lambda} (t) - A \left( x (t) + \frac{t}{\alpha} \hat{x} (t) \right) + b = 0
\end{cases}

(3.17)

\]

where \( \left( \nabla f (x (t)) + \beta A^T (Ax (t) - b) + A^T \left( \lambda (t) + \frac{t}{\alpha} \hat{\lambda} (t) \right) \right) \) is the \( i \)-th element of \( \nabla f (x (t)) + \beta A^T (Ax (t) - b) + A^T \left( \lambda (t) + \frac{t}{\alpha} \hat{\lambda} (t) \right) \).

Case 4: If \( \mathcal{X} \) is a closed and convex set, and its projection operator \( P_\mathcal{X} \) has a closed form solution, moreover, setting \( \psi (x) \) be \( \frac{1}{2} ||x||^2 + I_X (x) \), one has \( \partial \psi (x) = x + \mathcal{N}_\mathcal{X} (x) \) and \( D_\psi (x, y) = \frac{1}{2} ||x - y||^2 \), \( \forall x, y \in \mathcal{X} \). Furthermore, one has \( \psi^* (u) = \frac{1}{2} \left( ||u||^2 - ||u - P_\mathcal{X} (u)||^2 \right), \nabla \psi^* (u) = P_\mathcal{X} (u) \). According to the above discussion, the APDMD (3.3) reduces to a new accelerated primal-dual projection dynamical (APDPD) approach:

\[

\begin{cases}
\dot{x} (t) = \frac{\alpha}{2} \left( P_\mathcal{X} (u (t)) - x (t) \right), \\
\dot{u} (t) = - \frac{1}{\alpha} \left( \nabla f (x (t)) + \beta A^T (Ax (t) - b) + A^T v (t) \right), \\
\dot{\lambda} (t) = \frac{\alpha}{2} (v (t) - \lambda (t)), \\
\dot{v} (t) = \frac{\alpha}{2} \left( AP_\mathcal{X} (u (t)) - b \right), \\
x (t_0) = x_0, u (t_0) = u_0 \text{ with } x_0 = P_\mathcal{X} (u_0) \in \mathcal{X}; \lambda (t_0) = \lambda_0, v (t_0) = 0.
\end{cases}

(3.18)

3.6. APDMD for DCCP (3.1) in the smooth case. In this subsection, inspired by the ADPDMD (3.4) and distributed consensus theorem, an accelerated distributed primal-dual mirror dynamical (ADPDMD) approach for DCCP (3.1) in the smooth case is proposed and discussed.
For DCCP (3.1) with smooth convex objective functions, the proposed ADPDD with $\alpha \geq 2$ is given by

\[
\begin{align*}
\dot{x}_i(t) &= \frac{\alpha}{t} (\nabla \psi_i^*(u_i(t)) - x_i(t)), \\
\dot{u}_i(t) &= -\frac{1}{\alpha} (\nabla f_i(x_i(t)) + \beta \sum_{j=1}^{n} a_{ij} (x_i(t) - x_j(t)) + \sum_{j=1}^{n} a_{ij} \left( \lambda_i(t) + \frac{1}{\alpha} \dot{\lambda}_i(t) - \lambda_j(t) - \frac{1}{\alpha} \dot{\lambda}_j(t) \right)), \\
\dot{\lambda}_i(t) &= \frac{\alpha}{t} (v_i(t) - \lambda_i(t)), \\
v_i(t) &= \frac{1}{\alpha} \sum_{j=1}^{n} a_{ij} \left( \nabla \psi_i^*(u_i(t)) - \nabla \psi_j^*(u_j(t)) \right), \\
x_i(t_0) &= x_{i,0}, \nabla \psi_i^*(u_i(t_0)) = x_{i,0}, \\
\lambda_i(t_0) &= \lambda_{i,0}, v_i(t_0) = v_{i,0}, i = 1, ..., n.
\end{align*}
\]

(3.19)

Next, we will illustrate the accelerated convergence rate of the ADPDD (3.20) by the Lyapunov analysis method.

**Theorem 3.7.** Suppose Assumption 3.2 holds and let $(x(t), \lambda(t))$ and $(x^*, \lambda^*)$ be a solution trajectory and an optimal solution for ADPDMD (3.20) and DCCP (3.1), respectively. Then for any $(x(t_0), u(t_0), \lambda(t_0), v(t_0)) \in X \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$, the following statements are true:

(I): The trajectories $x(t)$ and $\lambda(t)$ of ADPDMD (3.20) are bounded for any $t \geq t_0 > 0$.

(II): Let $L_{\beta}(x(t), \lambda(t)) = f(x(t)) + \frac{2}{3} x(t)^T Lx(t) + \lambda(t)^T Lx(t)$, one has

\[
\begin{align*}
L_{\beta}(x(t), \lambda^*) - L_{\beta}(x^*, \lambda(t)) &\leq \frac{\alpha^2 V(t_0)}{t^2}, \\
x(t)^T Lx(t) &\leq \frac{2\alpha^2 V(t_0)}{\beta t^2}, \\
\int_{t_0}^{+\infty} t (L_{\beta}(x(t), \lambda^*) - L_{\beta}(x^*, \lambda(t))) \, dt &< +\infty, \\
\int_{t_0}^{+\infty} t \|Ax(t) - b\|^2 \, dt &< +\infty.
\end{align*}
\]

(3.21a) (3.21b) (3.21c) (3.21d)
(III): Let \( \lambda^* = \begin{cases} 0, & \text{if } L x(t) = 0, \\ \frac{x(t)}{\sqrt{x(t)^T L x(t)}}, & \text{if } L x(t) \neq 0, \end{cases} \). For any \( t \geq t_0 > 0 \), it follows that

\[
\begin{align*}
0 & \leq f(x(t)) - f(x^*) + \sqrt{x(t)^T L x(t)} \leq \frac{\alpha^2 V(t_0)}{t^2}, \\
-\frac{\alpha \sqrt{2} V(t_0)}{t \sqrt{\beta}} & \leq f(x(t)) - f(x^*) \leq \frac{\alpha^2 V(t_0)}{t^2}.
\end{align*}
\]

Proof. Consider the following candidate Lyapunov function:

\[
V(t) = V_1(t) + V_2(t) + V_3(t),
\]

with

\[
\begin{align*}
V_1(t) &= \frac{\alpha^2}{2} \left( L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t)) \right) \\
&= \frac{\alpha^2}{2} \left( f(x(t)) - f(x^*) + \frac{\beta}{2} x(t)^T L x(t) + L \lambda^* \right), \\
V_2(t) &= D_{\psi^*} (u(t), u^*) = \psi^* (u(t)) - \psi^* (u^*) - \nabla \psi^* (u^*)^T (u(t) - u^*), \\
V_3(t) &= D_h (v(t), \lambda^*) = \frac{1}{2} \| v(t) - \lambda^* \|^2,
\end{align*}
\]

Following the similar steps as the proof of Theorem 3.6, the proof can be obtained easily. Due to space limitations, we omit the proof. \( \square \)

3.7. APDMD for DEMO (3.2) in smooth case. In this subsection, to address the DEMO (3.2), where its objective function is smooth and convex, an accelerated distributed mirror dynamical (ADMD) approach is investigated based on APDMD (3.3). Recalling the DEMO (3.2), the coupled equations \( \sum_{i=1}^n A_i x_i = \sum_{i=1}^n d_i \in \mathbb{R}^m \) can be equivalently decomposed as \( \bar{x} - d + Ly = 0 \in \mathbb{R}^m \) with \( L = L_n \otimes I_m \) and an auxiliary variable \( y = \text{col}(y_1, ..., y_n) \in \mathbb{R}^{\sum_{i=1}^n p_i} \) by using the properties of the Laplacian matrix of undirected \( G \) (i.e., \( \ker(L_n) = \{\zeta | \zeta \in \mathbb{R} \}, \quad \text{rang}(L_n) = \{\omega \in \mathbb{R}^n | \omega^T 1 = 0 \} \}). Therefore, the DEMO (3.2) is equivalent to:

\[
\begin{align*}
\min f(x) &= \sum_{i=1}^n f_i(x_i), \\
\text{s.t. } \bar{x} - d + Ly = 0, x &\in \mathcal{X} = \prod_{i=1}^n \mathcal{X}_i,
\end{align*}
\]

where \( \bar{x} = \text{bdig}\{A_{p_1}, A_{p_2}, ..., A_{p_n}\} \in \mathbb{R}^{nm \times \sum_{i=1}^n p_i} \), \( d = \text{col}(d_1, d_2, ..., d_n) \in \mathbb{R}^{nm} \), and \( \mathcal{X} = \prod_{i=1}^n \mathcal{X}_i \) is the Cartesian product of set \( \mathcal{X}_i, i = 1, ..., n \).

To deal with the modified DEMO (3.24) with smooth convex objective functions, we propose the following accelerated distributed mirror dynamical (ADMD) approach with \( t \geq t_0 > 0 \) and \( \alpha \geq 2 \) as follows:

\[
\begin{align*}
\dot{x}(t) &= \frac{\alpha}{2} \left( \nabla \psi^* (u(t)) - x(t) \right), \\
\dot{\lambda}(t) &= \frac{\alpha}{2} \left( v(t) - \lambda(t) \right), \\
\dot{y}(t) &= \frac{\alpha}{2} \left( z(t) - y(t) \right), \\
x(t_0) &= x_0, u(t_0) = u_0 \quad \text{with } \nabla \psi^* (u_0) = x_0 \in \mathcal{X}, \\
\lambda(t_0) &= \lambda_0, v(t_0) = v_0, y(t_0) = y_0, z(t_0) = z_0.
\end{align*}
\]
Theorem 3.8. Suppose Assumption 3.2 holds and let \((x(t), \lambda(t), y(t), u(t), v(t), z(t))\) be a solution trajectory and \((x^*, \lambda^*, y^*, u^*, v^*, z^*)\) be an optimal solution of ADMD (3.25), respectively. Then, for any \((x(t_0), u(t_0), \lambda(t_0), v(t_0), y(t_0), z(t_0))\) \(\in\mathcal{X} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \), we have the following two statements:

(I): The trajectories \(x(t), \lambda(t)\) and \(y(t)\) of ADMD (3.25) are bounded for any \(t \geq t_0 > 0\).

(II): Let \(L\) with \(\beta = 1\), i.e., \(L(x(t), \lambda(t), y(t)) = f(x(t)) - \frac{1}{2}\lambda(t)^T L\lambda(t) + \lambda(t)^T (Ax(t) - d - Ly(t))\), it follows that

\[
(3.26a) \quad \|x(t) - y^*\|_2 \leq \frac{\alpha^2 E(t_0)}{t^2}, \quad \lambda(t)^T L\lambda(t) \leq \frac{2\alpha^2 E(t_0)}{t^2},
\]

\[
(3.26b) \quad \int_{t_0}^{\infty} t (\|x(t) - y^*\|_2 - \|x(t), \lambda(t)\|_2) \, dt < +\infty.
\]

Proof. (I): Designing a Lyapunov function as follows:

\[
(3.27) \quad E(t) = E_1(t) + E_2(t) + E_3(t),
\]

\[
\begin{align*}
E_1(t) &= \frac{\alpha}{2} (\|x(t) - y^*\|_2 - \|x^*, \lambda(t)\|_2), \\
E_2(t) &= \frac{\alpha}{2} \left( f(x(t)) - f(x^*) + (\lambda(t))^T (A(t) - d - \lambda(t)) \right), \\
E_3(t) &= \frac{1}{2} \|v(t) - \lambda(t)\|_2^2, \\
E_4(t) &= \frac{1}{2} \|z(t) - y^*\|_2^2.
\end{align*}
\]

Using the same proof steps as Theorem 3.6, we can get

\[
\dot{E}(t) = \dot{E}_1(t) + \dot{E}_2(t) + \dot{E}_3(t) + \dot{E}_4(t)
\]

\[
\leq - \frac{(\alpha - 2)t}{\alpha^2} \left( f(x(t)) - f(x^*) + (\lambda(t))^T A(x(t) - x^*) + \frac{1}{2} \lambda(t)^T L\lambda(t) \right)
\]

\[
= - \frac{(\alpha - 2)t}{\alpha^2} (\|x(t) - y^*\|_2 - \|x^*, \lambda(t)\|_2) \leq 0.
\]

Recall that the Lyapunov function \(E(t)\) is radically unbounded and positive with any \(t \geq t_0 > 0\). This implies that the trajectories of \(x(t), \lambda(t)\) and \(y(t)\) are bounded for any \(t \geq t_0 > 0\). The proof of (I) is therefore completed.

(II): Note that \(\dot{E}(t) \leq 0\), i.e., \(E(t)\) is nonincreasing on \([t_0, +\infty)\). Thus, for any \(t \geq t_0 > 0\), one has \(\frac{\alpha^2}{\alpha^2} (\|x(t) - y^*\|_2 - \|x^*, \lambda(t)\|_2) = E_3(t) \leq E(t) \leq E(t_0)\), i.e., \(\|x(t) - y^*\|_2 - \|x^*, \lambda(t)\|_2 \leq \frac{\alpha^2 E(t_0)}{t^2}\) holds. Furthermore, we have \(\lambda(t)^T L\lambda(t) \leq \frac{2\alpha^2 E(t_0)}{t^2}\). Thus, the conclusion of (3.26a) holds.

Since \(\dot{E}(t) \leq - \frac{(\alpha - 2)t}{\alpha^2} (\|x(t) - y^*\|_2 - \|x^*, \lambda(t)\|_2)\), thus, integrating it from \(t_0\) to \(+\infty\) and rearranging it, we have

\[
\int_{t_0}^{\infty} t (\|x(t) - y^*\|_2 - \|x^*, \lambda(t)\|_2) \, dt = \frac{\alpha^2}{\alpha - 2} (E(t_0) - E(+\infty)) \leq \frac{\alpha^2}{\alpha - 2} E(t_0) < +\infty.
\]

Therefore, the conclusion of (3.26b) holds. 
\[\square\]
4. Optimization approaches for problem (1.1) in the nonsmooth case.

For a vast amount of applications (e.g., signal processing, image processing, machine learning), nonsmooth functions are prevalent. To encompass these practical situations, we need to consider the case that the objective function $f(x)$ is nonsmooth. In order to adapt the APDMD (3.3) to this nonsmooth case, we will consider a corresponding smoothing accelerated primal-dual mirror dynamical (SAPDMD) approaches based on smoothing approximation in Subsection 2.6 as follows:

$$
\begin{aligned}
\dot{x}^\mu (t) &= \frac{\alpha}{t} (\nabla \psi^* (w^\mu (t)) - x^\mu (t)), \\
\dot{w}^\mu (t) &= -\frac{1}{\alpha} \left( \nabla_x \hat{f} (x^\mu (t), \mu (t)) + \beta A^T (Ax^\mu (t) - b) + A^T v^\mu (t) \right), \\
\dot{\lambda}^\mu (t) &= \frac{\alpha}{t} (w^\mu (t) - \lambda^\mu (t)), \\
\dot{v}^\mu (t) &= \frac{1}{\alpha} (A \nabla \psi^* (w^\mu (t)) - b (t)), \\
x^\mu (t_0) &= x^0, w^\mu (t_0) = u^0, \nabla \psi^* (u^0) = x^0, \\
\lambda^\mu (t_0) &= \hat{\lambda}^0, v^\mu (t_0) = v^0, \mu (t_0) = \mu_0,
\end{aligned}
$$

where $t \geq t_0 > 0$, $\beta > 0$, $\alpha \geq 2$ and $\mu (t) \leq \mu_0 \frac{t}{\alpha t_0}$. The superscript $\mu$ means the trajectories that are obtained by the smoothing dynamical approaches, which is used to distinguish the trajectories obtained by the dynamical approaches without smoothing approximation, i.e., APDMD (3.3), ADPDMD (3.20) and ADMD (3.25). The $\nabla_x \hat{f} (x^\mu (t), \mu (t))$ is the gradient of smoothing function of $\hat{f} (x^\mu (t), \mu (t))$ with respect of $x^\mu (t)$. The SAPDMD (4.1) is illustrated in Figure 3 (left).

Similar to APDMD (3.3), the SAPDMD (4.1) can be equivalent to the following smoothing second-order dynamical approach:

$$
\begin{aligned}
\ddot{x}^\mu (t) + \frac{\alpha+1}{t} \dot{x}^\mu (t) + \nabla^2 \psi^* (\nabla \psi (x^\mu (t) + \frac{\alpha}{\alpha} \dot{x}^\mu (t))) \left( \nabla_x \hat{f} (x^\mu (t), \mu (t)) + \beta A^T (Ax^\mu (t) - b) + A^T \lambda^\mu (t) + \frac{\alpha}{\alpha} \dot{\lambda}^\mu (t) \right) &= 0, \\
\ddot{\lambda}^\mu (t) + \frac{\alpha+1}{t} \dot{\lambda}^\mu (t) - A (x^\mu (t) + \frac{\alpha}{\alpha} \dot{x}^\mu (t)) + b &= 0, \\
x^\mu (t_0) &= x^0 \in \mathcal{X}, \dot{x}^\mu (t_0) = \dot{x}^0, \lambda^\mu (t_0) = \lambda^0, \dot{\lambda}^\mu (t_0) = \dot{\lambda}^0.
\end{aligned}
$$

![Fig. 3. (left) The demonstration of SAPDMD (4.1); (right) The demonstration of the proof of feasibility of SAPDMD (4.1).](image-url)
4.1. Nonsmooth dynamical approaches comparison. To solve nonsmooth optimization problems, many technologies have been studied and presented in designing dynamical optimized methods, such as, differential inclusion method [20, 3, 27], Moreau-Yosida regularization method [8, 4] and directional derivative method [47, 24].

Compared with the dynamical approaches based on techniques mentioned above for solving nonsmooth optimization problems, our proposed SAPDMD (4.1) based on smoothing approximation in Subsection 2.6 has the following differences and preponderances:

• The solutions may be different. The SAPDMD (4.1) has a strong global solution, which is similar to that in Definition 3.4. In [3, 20], some inertial differential inclusion dynamical approaches are investigated, but they have shock solutions. The dynamical differential inclusion approaches proposed in [27, 51] have a Carathéodory’s solution or Filippov’s solution.

• The SAPDMD (4.1) can achieve the optimal solution of problem (1.1) in the nonsmooth case when \( \mu \to 0 \), which is different from the works in [41, 56]. In the famous work [41], Nesterov’s pioneered a smoothing method based on the Fenchel conjugate technique to deal with nonsmooth optimization problems and applied it to design accelerated algorithms with a \( \varepsilon \)-solution, i.e., \( f(x^\varepsilon) - \min f(x) \leq \varepsilon \), where \( \varepsilon \) is a small constant, but it’s not equal to 0.

• The proposed SAPDMD (4.1) does not require solving some subproblems. However, the accelerated dynamic approaches in [8, 4] need to use the Moreau-Yosida approximation \( f_\rho(x) = \min_{x \in A} \left\{ f(\xi) + \frac{1}{2\rho} \| x - \xi \|_2^2 \right\} \). the accelerated dynamic approaches [47, 24] need to utilize \( d(x; \dot{x}) = \arg \max_{f \in \partial f(x)} f^T \dot{x} \), for them in general, there are no closed formulas available. This is undesirable from the point of view of numerical calculation.

• The existence and uniqueness of the global solutions for SAPDMD (4.1) can be easily guaranteed by the Cauchy-Lipschitz-Picard theorem. However, the uniqueness of the solutions in directional derivative method [27] and differential inclusion method [47, 24] may not be guaranteed.

4.2. Feasibility, existence and uniqueness of strong global solution for SAPDMD. In this subsection, we illustrate the feasibility, existence and uniqueness of a strong global solution \( x(t) \) for the SAPDMD (4.1) by the same way as in Subsection 3.3.

**Lemma 4.1.** For any \( (x^\mu(t_0), u^\mu(t_0), \lambda^\mu(t_0), \nu^\mu(t_0)) \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \) with \( \nabla \psi^\mu(x^\mu(t_0)) = x^\mu(t_0) \), then the variable \( x^\mu(t) \) of SAPDMD is always in \( \mathcal{X} \), \( \forall \ t \geq t_0 > 0 \), i.e., the feasibility of \( x(t) \) is satisfied.

**Proof.** The proof follows the same arguments as in Lemma 3.3 by the reductio and separation hyperplane theorem, we omit it here due to the limitation of space.

**Theorem 4.2.** For any initial values \( (x^\mu(t_0), u^\mu(t_0), \lambda^\mu(t_0), \nu^\mu(t_0)) \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \), there exists a unique strong global solution of SAPDMD (4.1).

**Proof.** Let \( Y^\mu(t) = (x^\mu(t), u^\mu(t), \lambda^\mu(t), \nu^\mu(t)) \), then the SAPDMD (4.1) is equivalent to

\[
\begin{align*}
\dot{Y}^\mu(t) &= H(t, Y^\mu(t)) \\
Y^\mu(t_0) &= (x^\mu(t_0), u^\mu(t_0), \lambda^\mu(t_0), \nu^\mu(t_0)),
\end{align*}
\]
where
\[
H(t, Y^\mu) = \left( \frac{\alpha}{t} (\nabla \psi^*(u^\mu(t)) - x^\mu(t)), -\frac{t}{\alpha} \left( \nabla_x f(x^\mu(t), \mu(t)) - A^T v^\mu(t) \right) - A^T v^\mu(t) \right).
\]

(4.4)

To prove the existence and uniqueness of the strong global solution \(Y^\mu(t)\) generated by SAPDMD (4.1) by the Cauchy-Lipschitz-Picard theorem [15], the following conditions need to be satisfied:

(I) For every \(t \in [t_0, +\infty)\), the mapping \(H(t, \cdot)\) is \(l(t)\)-Lipschitz continuous and \(l(\cdot) \in L^1_{\text{loc}}([t_0, +\infty))\).

(II) For any \(Y^\mu \in X \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m\), we have \(H(\cdot, Y^\mu) \in L^1_{\text{loc}}([t_0, +\infty), X \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m)\).

The proof of (I). Let \(t \in [t_0, +\infty)\) be fixed and use the Lipschitz continuous of \(\nabla \psi^*, \nabla_x f(x, \mu)\). Then, for any \(Y^\mu, \hat{Y}^\mu\), we have
\[
\left\| F(t, Y^\mu(t)) - F(t, \hat{Y}^\mu(t)) \right\| \leq \left( \frac{t^2}{\alpha^2} \left( 4\beta^2 \delta_{\text{max}} (A^T A)^2 + \frac{4\ell^2}{\mu^2(t)} \right) + \delta_{\text{max}} (A^T A) \left\| \nabla \psi^* (L_\psi) \right\| \right)^\frac{1}{2} \left\| Y^\mu(t) - \hat{Y}^\mu(t) \right\|.
\]

By using the notation \(l(t) = \left( \frac{t^2}{\alpha^2} \left( 4\beta^2 \delta_{\text{max}} (A^T A)^2 + \frac{4\ell^2}{\mu^2(t)} + \delta_{\text{max}} (A^T A) \right) \right)\), one has \(\left\| H(t, Y^\mu(t)) - H(t, \hat{Y}^\mu(t)) \right\| \leq l(t) \left\| Y^\mu(t) - \hat{Y}^\mu(t) \right\|\).

Note that \(l(t)\) is continuous on \([t_0, +\infty)\). Hence \(l(\cdot)\) is integrable on \([t_0, T]\) for all \(t_0 < T < +\infty\).

The proof of (II). Let \(Y \in X \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m\) and \(t_0 < T < +\infty\), it holds that
\[
\int_{t_0}^T \left\| H(t, Y^\mu(t)) \right\| dt \leq \left( \frac{8\alpha^2}{t^4} + \frac{2\ell^2}{\alpha^2} \left( 4\delta_{\text{max}} (A^T A) + 4\beta^2 \delta_{\text{max}} (A^T A)^2 + 2 \right) \right)^\frac{1}{2},
\]

and the conclusion holds by employing the continuity of the function
\[
t \to \left( \frac{8\alpha^2}{t^4} + \frac{2\ell^2}{\alpha^2} \left( 4\delta_{\text{max}} (A^T A) + 4\beta^2 \delta_{\text{max}} (A^T A)^2 + 2 \right) \right)^\frac{1}{2} \text{ on } [t_0, T].
\]

The existence and uniqueness of \(Y^\mu(t)\) to the dynamical system (4.4) can be guaranteed by the Cauchy-Lipschitz-Picard theorem, consequently, the existence and uniqueness of the trajectories of SAPDMD (4.1) also hold. □
4.3. The accelerated convergence of the SAPDMD. In this subsection, we will illustrate the accelerated convergence properties of the SAPDMD (4.1) based on the Lyapunov analysis method.

A natural question is whether the Lyapunov analysis method in the smooth case is still effective in the nonsmooth case. The answer is affirmative, provided some care is taken in the main three steps of our analysis. First, a time-dependent parameter \( \mu (t) \) needs to be introduced in the Lyapunov function. Second, when taking the time derivative of Lyapunov function, it requires utilizing the full differentiation of the smoothing approximation function, i.e., \( \nabla_{\mu} f (x (t), \mu (t)) \) needs to be considered. The third factor is boundedness of gradient and Lipschitz continuous of \( \mu \). In turn, all the results and estimation we have presented in the previous sections can be transferred to this more general nonsmooth context.

**Theorem 4.3.** Suppose that Assumption 3.1 holds and the objective function is nonsmooth convex. Let \( (x (t), y (t)) \) and \( (x^*, y^*) \) be a solution trajectory and an optimal solution of SAPDMD (4.1), respectively. Defining a Lyapunov function as 
\[
\hat{L}_\beta (x^\mu (t), \lambda^\mu (t)) = \hat{f} (x^\mu (t), \mu (t)) + \lambda^\mu (t)^T (Ax^\mu (t) - b) + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2
\]
with \( \beta \geq 0 \), then, for any \( (x^\mu (t), \lambda^\mu (t)) \in \mathcal{X} \times \mathbb{R}^m \), we have

(I): The trajectories of \( x^\mu (t) \) and \( \lambda^\mu (t) \) are bounded for any \( t \geq t_0 > 0 \).

(II): Then, we have

\[
0 \leq \hat{\mathcal{L}}_{\beta} (x^\mu (t), \lambda^\mu (t)) - \hat{\mathcal{L}}_{\beta} (x^*, \lambda^\mu (t)) + 4\kappa_j \mu (t) \leq \frac{\alpha^2 V (t_0)}{t^2},
\]

\[
\| Ax^\mu (t) - b \|^2 \leq \frac{2\alpha^2 V (t_0)}{\beta t^2},
\]

\[
\int_{t_0}^{+\infty} \| Ax^\mu (t) - b \|^2 dt \leq +\infty,
\]

\[
\int_{t_0}^{+\infty} t \left( \hat{\mathcal{L}}_{\beta} (x^\mu (t), \lambda^\mu (t)) - \hat{\mathcal{L}}_{\beta} (x^*, \lambda^\mu (t)) + 2\kappa_j \mu (t) \right) dt \leq +\infty.
\]

(III): Let \( \lambda^* = \begin{cases} 0, & \text{if } Ax^\mu (t) - b = 0, \\ \frac{Ax^\mu (t) - b}{\| Ax^\mu (t) - b \|}, & \text{if } Ax^\mu (t) - b \neq 0, \end{cases} \) and for any \( t \geq t_0 > 0 \), it follows that

\[
0 \leq f (x^\mu (t)) - f (x^*) + \| Ax^\mu (t) - b \| \leq \frac{\alpha^2 V (t_0)}{t^2};
\]

\[
-\frac{\alpha \sqrt{2} V (t_0)}{t \sqrt{\beta}} \leq f (x^\mu (t)) - f (x^*) \leq \frac{\alpha^2 V (t_0)}{t^2}.
\]

**Proof.** (I): Consider a well-designed smooth Lyapunov function

\[
V (t) = V_1 (t) + V_2 (t) + V_3 (t),
\]

with

\[
\begin{align*}
V_1 (t) &= \frac{t^2}{\alpha^2} \left( \hat{\mathcal{L}}_{\beta} (x^\mu (t), \lambda^* (t)) - \hat{\mathcal{L}}_{\beta} (x^*, \lambda^\mu (t)) + 4\kappa_j \mu (t) \right) \\
&= \frac{t^2}{\alpha^2} \left( \hat{f} (x^\mu (t), \mu (t)) - f (x^*, \mu (t)) + 4\kappa_j \mu (t) \\
&\quad + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + (\lambda^*)^T A (x^\mu (t) - x^*) \right) \\
V_2 (t) &= D_{\lambda^*} (v^\mu (t), u^*) - \psi^* (u^\mu (t)) - \psi^* (u^*) - \nabla \psi^* (u^*)^T (u^\mu (t) - u^*), \\
V_3 (t) &= D_{\lambda^*} (V^\mu (t), \lambda^*) = \frac{1}{2} \| V^\mu (t) - \lambda^* \|^2.
\end{align*}
\]
Note that the function $V_1(t)$ is positive for any $t \geq t_0 > 0$ since
\[
\hat{L}_\beta (x^\mu (t), \lambda^*) - \hat{L}_\beta (x^*, \lambda^* (t)) + 4\kappa_j \mu (t)
\geq f(x^\mu (t)) - f(x^*) + (\lambda^*)^T A (x^\mu (t) - x^*) + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + 2\kappa_j \mu (t)
\]
\[
= L_\beta (x^\mu (t), \lambda^*) - L_\beta (x^*, \lambda^* (t)) + 2\kappa j \mu (t) \geq 0,
\]
where the first equality holds from $\hat{f}(x^\mu (t), \mu (t)) - f(x^\mu (t)) + \kappa_j \mu (t) \geq 0$, $f(x^*) - f(x^*, \mu (t)) + \kappa_j \mu (t) \geq 0$ in Definition 2.6 (iv), and the second inequality is satisfied due to (2.2) in Subsection 2.2. In addition, since $\psi^* (\cdot)$ and $h (\cdot) = \frac{1}{2} \| \cdot \|^2$ are convex, such that $D\psi^* (u^\mu (t), u^*) \geq 0$, $D\psi^* (u^\mu (t), \lambda^*) \geq 0$. Thus, we end up with $V(t) \geq 0$.

The derivatives of $V_1(t)$, $V_2(t)$ and $V_3(t)$ along the trajectory of SAPDMD (4.1) satisfy
\[
\dot{V}_1(t) = \frac{2t}{\alpha^2} \left( \hat{f}(x^\mu (t), \mu (t)) - f(x^*, \mu (t)) + (\lambda^*)^T A (x^\mu (t) - x^*) + 4\kappa_j \mu (t) \right)
+ \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2
\times (\nabla_x \hat{f}(x^\mu (t), \mu (t)))^T + \frac{t}{\alpha} (\nabla \psi^* (u^\mu (t)) - \nabla \psi^* (u^*))^T
\times (\nabla_x \hat{f}(x^\mu (t), \mu (t)))^T + \frac{2t^2}{\alpha^2} \kappa_j \mu (t)
+ \frac{t^2}{\alpha^2} \left( \nabla_x \hat{f}(x^\mu (t), \mu (t)) \right) + \nabla_{\mu} \hat{f}(x^* (t), \mu (t)) + 2\kappa_j \mu (t),
\]
\[
\dot{V}_2(t) = - \frac{t}{\alpha} (\nabla \psi^* (u^\mu (t)) - \nabla \psi^* (u^*))^T (A^T u^\mu (t) - A^T \lambda^*)
- (\nabla_x \hat{f}(x^\mu (t), \mu (t)))^T + \beta A^T (x^\mu (t) - x^*) + A^T \lambda^*)^T
\times \frac{t}{\alpha} (\nabla \psi^* (u^\mu (t)) - \nabla \psi^* (u^*))
\]
\[
\dot{V}_3(t) = (v^\mu (t) - \lambda^*)^T v^\mu (t) = \frac{t}{\alpha} (v^\mu (t) - \lambda^*)^T A (\nabla \psi^* (u^\mu (t)) - x^*).
\]

Combining (4.8), (4.9), (4.10) and rearranging it yields
\[
\dot{V}(t) = \dot{V}_1(t) + \dot{V}_2(t) + \dot{V}_3(t)
\leq \frac{2t}{\alpha^2} \left( \hat{f}(x^\mu (t), \mu (t)) - \hat{f}(x^*, \mu (t)) + \beta \| Ax^\mu (t) - b \|^2 + 4\kappa_j \mu (t)
+ (\lambda^*)^T A (x^\mu (t) - x^*) \right) - \frac{t}{\alpha} \left( \hat{f}(x^\mu (t), \mu (t)) - \hat{f}(x^*, \mu (t)) \right)
+ \beta \| Ax^\mu (t) - b \|^2 + (\lambda^*)^T A (x^\mu (t) - x^*) + \frac{2t^2}{\alpha^2} \kappa_j \mu (t)
\leq \frac{(2 - \alpha)}{2^2} \left( \hat{f}(x^\mu (t), \mu (t)) - \hat{f}(x^*, \mu (t)) \right) + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + 4\kappa_j \mu (t)
+ (\lambda^*)^T A (x^\mu (t) - x^*) + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2
\]
\[
+ \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + 4\kappa_j \mu (t)
\]
\[
+ (\lambda^*)^T A (x^\mu (t) - x^*) + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + 4\kappa_j \mu (t)
\]
\[
+ (\lambda^*)^T A (x^\mu (t) - x^*) + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + 4\kappa_j \mu (t)
\]
\[
+ (\lambda^*)^T A (x^\mu (t) - x^*) + \frac{\beta}{2} \| Ax^\mu (t) - b \|^2 + 4\kappa_j \mu (t)
\]
where the first inequality holds since \( \nabla_{\bar{\mu}} \hat{f}(x^\mu(t), \mu(t)) + \nabla_{\bar{\mu}} \hat{f}(x^\mu(t), \mu(t)) + 2\kappa_{\bar{f}} \geq 0 \) and \( \hat{\mu}(t) \leq 0 \), the second inequality is satisfied because of the convexity of \( \hat{f}(x^\mu(t), \mu(t)) \) of \( x^\mu(t) \in \mathcal{X} \) with any fixed \( \mu(t) \in [0, \bar{\mu}] \), the third inequality holds from \( \mu(t) \leq \frac{1}{2\alpha} \) (i.e., \( \hat{\mu} \leq -\mu_0(2\alpha) t^{(-2\alpha - 1)} \) and \( \frac{\beta}{2\alpha} \kappa_{\bar{f}} \geq 0 \) imply \( \frac{2\beta}{\alpha} \kappa_{\bar{f}} \mu(t) + \frac{2\beta}{\alpha} \kappa_{\bar{f}} \hat{\mu}(t) \leq \frac{2\beta}{\alpha} \kappa_{\bar{f}} \mu_0 t^{-2\alpha} - \frac{2\beta}{\alpha} \kappa_{\bar{f}} \mu_0 (2\alpha) t^{-2\alpha - 1} = \frac{2\beta}{\alpha} \kappa_{\bar{f}} \mu_0 t^{-2\alpha} - \frac{2\beta}{\alpha} \kappa_{\bar{f}} \mu_0 t^{-2\alpha} = 0 \), and the last inequality holds since \( \alpha \geq 2 \) and \( \hat{\mathcal{L}}_{\beta}(x^\mu(t), \lambda^\mu(t)) - \hat{\mathcal{L}}_{\beta}(x^\mu(t), \lambda^\mu(t)) + 4\kappa_{\bar{f}} \mu(t) \geq 0 \).

Recall that the Lyapunov function \( V(t) \) is radically unbounded and positive with any \( t \geq t_0 > 0 \). This implies that the trajectories of \( x^\mu(t) \) and \( \lambda^\mu(t) \) are bounded for any \( t \geq t_0 > 0 \). The proof of (I) is therefore completed.

(II): From (4.11), we have that \( \dot{V}(t) \leq 0 \), i.e., \( V(t) \) is nonincreasing on \( t \in [t_0, +\infty) \). It yields to \( \frac{\alpha - 2}{\alpha^2} \int_{t_0}^{+\infty} t \left( \hat{\mathcal{L}}_{\beta}(x^\mu(t), \lambda^\mu(t)) - \hat{\mathcal{L}}_{\beta}(x^\mu(t), \lambda^\mu(t)) + 4\kappa_{\bar{f}} \mu(t) \right) dt = V(t_0) - V(+\infty) \leq V(t_0) < +\infty \),\n
\[
\beta \int_{t_0}^{+\infty} t \|Ax^\mu(t) - b\|^2 dt = V(t_0) - V(+\infty) \leq V(t_0) < +\infty,
\]

which implies both (4.5c) and (4.5d) hold.

(III): The equations \( \lambda^\mu = \begin{cases} 0, & \text{if } Ax^\mu(t) - b = 0, \\ \frac{Ax^\mu(t) - b}{\|Ax^\mu(t) - b\|}, & \text{if } Ax^\mu(t) - b \neq 0 \end{cases} \), and (4.5a) give

\[
\|Ax^\mu(t) - b\| \leq \frac{\alpha \sqrt{2V(t_0)}}{\sqrt{\beta}}.
\]

According to the upper-bound of (2.2) in Subsection 2.2 with \( \beta = 0 \), one has

\[
\mathcal{L}_0(x^\mu(t), \lambda^\mu(t)) = f(x^\mu(t)) - f(x^*) \geq -(Ax^\mu(t) - b)^T \lambda^\mu(t), \forall (x^\mu(t), \lambda^\mu(t)) \in \mathcal{X} \times \mathbb{R}^m,
\]

and from (iv) in Definition 2.6, one also has

\[
\hat{f}(x^\mu(t), \mu(t)) \geq f(x^\mu(t)) - f(x^*) \geq -(Ax^\mu(t) - b)^T \lambda^\mu(t) \geq -\|Ax^\mu(t) - b\| \geq -\frac{\alpha \sqrt{2V(t_0)}}{t\sqrt{\beta}}, \forall (x^\mu(t), \lambda^\mu(t)) \in \mathcal{X} \times \mathbb{R}^m.
\]
Since \( \|Ax^\mu(t) - b\| \geq 0 \) and (4.5a) imply \( f(x^\mu(t)) - f(x^*) \leq \frac{\alpha^2V(t_0)}{t^2} \), further, we deduce for any \( t \geq t_0 > 0 \) that \( -\frac{\alpha^2V(t_0)}{t^2 \sqrt{\beta}} \leq f(x^\mu(t)) - f(x^*) \leq \frac{\alpha^2V(t_0)}{t^2} \) is true. Therefore, we can obtain the conclusion (III).

Similar to the smooth case, the SAPDMD (4.1) can also be used to address the DCCP (3.1) and DEMO (3.2) with nonsmooth convex objections.

4.4. SAPDMD for DCCP in the nonsmooth case. Based on the SAPDMD (4.1), a smoothing accelerated distributed primal-dual mirror dynamical (SADPDMD) approach for DCCP (3.1) with nonsmooth convex objective functions is given by:

\[
\begin{align*}
    \dot{x}^\mu(t) &= \frac{\alpha}{\tau} \left( \nabla \psi^* (u^\mu(t)) - x^\mu(t) \right), \\
    \dot{u}^\mu(t) &= -\frac{1}{\alpha} \left( \nabla \hat{f} (x^\mu(t), \mu(t)) + \beta Lx^\mu(t) + L(v^\mu(t)) \right), \\
    \dot{\lambda}^\mu(t) &= \frac{\beta}{\tau} (v^\mu(t) - \lambda^\mu(t)), \\
    \dot{v}^\mu(t) &= \frac{\alpha}{\tau} L \nabla \psi^* (u^\mu(t)), \\
    x^\mu(t_0) &= x^0, \lambda^\mu(t_0) = \lambda^0, v^\mu(t_0) = v^0,
\end{align*}
\]

where \( t \geq t_0 > 0, \beta > 0, \alpha \geq 2, \) and \( \mu(t) \leq \frac{\mu_0}{\tau \sqrt{\beta}}. \)

4.5. The accelerated convergence of the SADPDMD. Now, let’s discuss accelerated convergence properties of SADPDMD (4.12) with the help of Lyapunov analysis tool.

**Theorem 4.4.** Suppose that Assumption 3.2 holds and the objective function is nonsmooth. Let \( (x^\mu(t), \lambda^\mu(t)) \) and \( (x^*, \lambda^*) \) be a solution trajectory and an optimal solution of SADPDMD (4.12), respectively. Let \( \hat{L}_\beta(x^\mu(t), \lambda^\mu(t)) = \hat{f}(x^\mu(t), \mu(t)) + \frac{\beta}{\tau} \left( x^\mu(t) \right)^T L \left( x^\mu(t) \right) + \left( \lambda^\mu(t) \right)^T Lx^\mu(t) \) with \( \beta \geq 0, \) then, for \( \alpha \geq 2, \) one has

(I): \( x^\mu(t), \lambda^\mu(t), u^\mu(t), v^\mu(t) \) are bounded for any \( t \geq t_0 > 0. \)

(II): Then, we have

\[
\begin{align*}
    0 &\leq \hat{L}_\beta(x^\mu(t), \lambda^* ) - \hat{L}_\beta(x^*, \lambda^\mu(t)) + 4\kappa \mu(t) \leq \frac{\alpha^2V(t_0)}{t^2}, \\
    x^\mu(t) Lx^\mu(t) &\leq \frac{2\alpha^2V(t_0)}{\beta t^2}, \\
    \int_{t_0}^{+\infty} x^\mu(t) Lx^\mu(t) dt &\leq +\infty, \\
    \int_{t_0}^{+\infty} t \left( \hat{L}_\beta(x^\mu(t), \lambda^* ) - \hat{L}_\beta(x^*, \lambda^\mu(t)) + 2\kappa \mu(t) \right) dt &< +\infty,
\end{align*}
\]

(III): Let \( \lambda^* = \begin{cases} 0, & \text{if } Lx^\mu(t) = 0, \\ \frac{Lx^\mu(t)}{\|Lx^\mu(t)\|}, & \text{if } Lx^\mu(t) \neq 0. \end{cases} \) For any \( t \geq t_0 > 0, \) it follows that

\[
\begin{align*}
    f(x^\mu(t)) - f(x^*) + \|Lx^\mu(t)\| &\leq \frac{\alpha^2V(t_0)}{t^2}; \\
    -\frac{\alpha \sqrt{2V(t_0)}}{t \sqrt{\beta}} &\leq f(x^\mu(t)) - f(x^*) \leq \frac{\alpha^2V(t_0)}{t^2}.
\end{align*}
\]

**Proof.** Construct the candidate smoothing Lyapunov function as

\[
\mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2 + \mathcal{V}_3,
\]
with
\[
\begin{align*}
\mathcal{V}_1(t) &= \frac{\beta}{2} \left( \bar{L}_\beta (x^\mu (t), \lambda^*) - \hat{L}_\beta (x^*, \lambda^*) + 4\kappa f \mu (t) \right), \\
\mathcal{V}_2(t) &= D_{\psi} (u^\mu (t), u^\mu) = \psi^* (u^\mu (t)) - \psi^* (u^*) - \nabla \psi^* (u^*)^T (u^\mu (t) - u^*), \\
\mathcal{V}_3(t) &= D_h (v^\mu (t), \lambda^*) = \frac{1}{2} \| v^\mu (t) - \lambda^* \|^2.
\end{align*}
\]

Following similar steps as the proof in Theorem 4.3, the conclusions can be obtained easily. Due to space limitations, we omit the proof here.

### 4.6. SADMD for DEMO in the nonsmooth case

To address the DEMO (3.2) where the objective function is nonsmooth, a smoothing accelerated distributed mirror dynamical (SADMD) approach inspired by the SAPDMD (4.1) and distributed consensus theorem is given by

\[
\begin{align*}
\dot{x}^\mu (t) &= \frac{\beta}{2} (\nabla \psi^* (u^\mu (t)) - x^\mu (t)), \\
\dot{u}^\mu (t) &= - \frac{\alpha}{\alpha} \left( \nabla_x \hat{f} (x^\mu (t), \mu (t)) + \bar{A}^T v^\mu (t) \right), \\
\dot{\lambda}^\mu (t) &= \frac{\beta}{2} (v^\mu (t) - \lambda^\mu (t)), \\
\dot{v}^\mu (t) &= \frac{1}{\alpha} (\bar{A} \psi^* (u^\mu (t)) - \bar{d} - \bar{L} \lambda^\mu (t) + \bar{L} \psi^* (u^\mu (t))), \\
\dot{y}^\mu (t) &= \frac{\beta}{\alpha} (z^\mu (t) - y^\mu (t)), \\
\dot{z}^\mu (t) &= - \frac{1}{\alpha} \bar{L} v^\mu (t), \\
x^\mu (t_0) &= x^\mu_0, u^\mu (t_0) = u^\mu_0, \text{ with } \nabla \psi^* (u^\mu (t_0)) = x^\mu_0, \lambda^\mu_0, v^\mu (t_0) = v^\mu_0, y^\mu (t_0) = y^\mu_0, z^\mu (t_0) = z^\mu_0,
\end{align*}
\]

where \( t \geq t_0 > 0, \beta > 0, \alpha \geq 2, \text{ and } \mu (t) \leq \frac{\mu_0}{1 + t}. \)

The accelerated convergence properties of SADMD (4.16) will be demonstrated in the following theorem.

**Theorem 4.5.** Suppose that Assumption 3.2 holds, except that the objective function is nonsmooth. Let \((x^\mu (t), \lambda^\mu (t), y^\mu (t))\) and \((x^*, \lambda^*, y^*)\) be a solution trajectory and an optimal solution of SADMD (4.16), respectively. Then, for any initial values \((x^\mu (t_0), u^\mu (t_0), \lambda^\mu (t_0), v^\mu (t_0), y^\mu (t_0), z^\mu (t_0)) \in \mathcal{X} \times \mathbb{R}^{\sum_{i=1}^p} \times \mathbb{R}^{\sum_{i=1}^p} \times \mathbb{R}^{\sum_{i=1}^p} \times \mathbb{R}^{\sum_{i=1}^p} \times \mathbb{R}^{\sum_{i=1}^p} \times \mathbb{R}^{\sum_{i=1}^p}, \) the following statements are true.

1. **(I):** The trajectories of \(x^\mu (t), \lambda^\mu (t)\) and \(y^\mu (t)\) are bounded for any \( t \geq t_0 > 0. \)
2. **(II):** Let
\[
\hat{L}_1 (x, \lambda, y) = \hat{f} (x^\mu (t), \mu (t)) - \frac{1}{2} \lambda^\mu (t)^T \bar{L} \lambda^\mu (t) + \lambda^\mu (t)^T \bar{A} x^\mu (t) - d - Ly^\mu (t),
\]
then, it follows
\[
\begin{align*}
\hat{L}_1 (x^\mu (t), y^*, \lambda^*) - \hat{L}_1 (x^*, y^*, \lambda^*) + 4\kappa f \mu (t) \leq \frac{\alpha t^2 \mathcal{E}(t_0)}{t^2}, \\
\lambda^\mu (y)^T \bar{L} \lambda^\mu (t) \leq \frac{2 \alpha t^2 \mathcal{E}(t_0)}{t^2}, \\
\int_{t_0}^{+\infty} t \left( \hat{L}_1 (x^\mu (t), y^*, \lambda^*) - \hat{L}_1 (x^*, y^*, \lambda^*(t)) + 4\kappa f \mu (t) \right) dt < +\infty.
\end{align*}
\]

**Proof.** (I): Consider a smoothing Lyapunov function as follows:

\[
\mathcal{E}(t) = \mathcal{E}_1 (t) + \mathcal{E}_2 (t) + \mathcal{E}_4 (t),
\]
with

$$
\begin{align*}
E_1(t) &= \frac{\alpha^2}{\alpha^2} \left( L_1(x^\mu(t), y^\lambda, \lambda^*) - L_1(x^*, y^*, \lambda^*) + 4\kappa f \mu(t) \right), \\
E_2(t) &= \frac{\alpha^2}{\alpha^2} \left( \dot{f}(x^\mu(t), \mu(t)) - \dot{f}(x^*, \mu(t)) + \frac{1}{2} \lambda^T L \lambda^\mu(t) \right), \\
E_3(t) &= \lambda^T (x^\mu(t), \lambda^*) = \frac{1}{2} \|v^\mu(t) - \lambda^*\|^2, \\
E_4(t) &= \lambda^T (z^\mu(t), y^*) = \frac{1}{2} \|z^\mu(t) - y^*\|^2, \\
E_5(t) &= 4 \kappa f \mu(t) + (\lambda^*)^T (\dot{A} x^\mu(t) - d - Ly^*)
\end{align*}
$$

where $D_{\psi^*}(u, u^*)$ and $D_h(v, \lambda^*)$ are the Bregman divergences associated with $\psi^*$ for $u, u^*$ and $v, \lambda^*$. The time derivative of $E(t)$ with $\nabla \psi^*(u^*) = x^*$ is

$$
\dot{E}(t) = \dot{E}_1(t) + \dot{E}_2(t) + \dot{E}_3(t) + \dot{E}_4(t)
$$

$$
\dot{E}(t) \leq \frac{2t}{\alpha^2} \left( \dot{f}(x^\mu(t), \mu(t)) - \dot{f}(x^*, \mu(t)) + (\lambda^*)^T \dot{A}(x^\mu(t) - x^*) + 4\kappa f \mu(t) \right)
$$

$$
+ \frac{1}{2} \lambda^T L \lambda^\mu(t) - \frac{t}{\alpha} (L \lambda^\mu(t))^T \lambda^\mu(t) + \frac{4t}{\alpha^2} \kappa f \mu(t) + \frac{2\mu t}{\alpha^2} \kappa f \mu(t)
$$

$$
\leq -\frac{(\alpha - 2)t}{\alpha^2} \left( \dot{f}(x^\mu(t), \mu(t)) - \dot{f}(x^*, \mu(t)) + (\lambda^*)^T \dot{A}(x^\mu(t) - x^*) \\
+ 4\kappa f \mu(t) + \frac{1}{2} \lambda^T L \lambda^\mu(t) \right)
$$

$$
= -\frac{(\alpha - 2)t}{\alpha^2} \left( \dot{L}_1(x^\mu(t), y^*, \lambda^*) - \dot{L}_1(x^*, y^*, \lambda^*(t)) + 4\kappa f \mu(t) \right) \leq 0,
$$

where the first inequality holds because of $\nabla \mu \dot{f}(x^\mu(t), \mu(t)) + \nabla \mu \dot{f}(x^*, \mu(t)) + 2\kappa f \geq 0$, the second inequality is satisfied due to the convexity of $\dot{f}(x^\mu(t), \mu(t))$ of $x^\mu(t) \in X$ with any fixed $\mu(t) \in [0, \mu]$ and the property of $\mu(t) \leq \frac{\mu_0}{t \alpha^2}$ (i.e., $\dot{\mu}(t) \leq -\mu_0(2\alpha)^t t^{-2\alpha}$), and the third inequality is established since $\alpha \geq 2$ and $\dot{L}_1(x^\mu(t), \lambda^*) - \dot{L}_1(x^*, \lambda^*(t)) + 4\kappa f \mu(t) \geq 0$.

Recall that the Lyapunov function $E(t)$ is radially unbounded and positive for any $t \geq t_0 > 0$. This implies that the trajectories of $x^\mu(t), \lambda^\mu(t)$ and $y^\mu(t)$ are bounded for any $t \geq t_0 > 0$. The proof of (I) is completed.

(II): From (4.19), we have that $\dot{E}(t) \leq 0$, i.e., $E(t)$ is nonincreasing on $[t_0, +\infty)$. Thus, for any $t \geq t_0 > 0$, one has $\dot{L}_1(x^\mu(t), y^*, \lambda^*) - \dot{L}_1(x^*, y^*, \lambda^*(t)) + 4\kappa f \mu(t) = E_1(t) \leq E(t) \leq E(t_0)$, i.e., $\dot{L}_1(x^\mu(t), y^*, \lambda^*) - \dot{L}_1(x^*, y^*, \lambda^*(t)) + 4\kappa f \mu(t) \leq \frac{\alpha^2 E(t_0)}{t^2}$.

$$
\text{(4.17a)}
$$

In addition, the (4.17a) implies $\lambda^\mu(t)^T L \lambda^\mu(t) \leq \frac{2\alpha^2 E(t_0)}{t^2}$, i.e., the (4.17b) holds.

Since $\dot{E}(t) \leq -\frac{(\alpha - 2)t}{\alpha^2} \left( \dot{L}_1(x^\mu(t), y^*, \lambda^*) - \dot{L}_1(x^*, y^*, \lambda^*(t)) + 4\kappa f \mu(t) \right)$, thus,
integrating it from \(t_0\) to \(+\infty\) and rearranging it, we have

\[
\int_{t_0}^{+\infty} t \left( \tilde{L}_1(x^*, y^*, \lambda^*) - \tilde{L}_1(x^*, y^*, \lambda^*(t)) + 4\kappa_f \mu(t) \right) dt \\
= \frac{\alpha^2}{\alpha - 2} (E(t_0) - E(+\infty)) \leq \frac{\alpha^2}{\alpha - 2} E(t_0) < +\infty,
\]

to complete the proof of (4.17d).

5. Numerical experiment. In this section, we give several numerical experiments to illustrate the effectiveness and superiority of the proposed dynamical approaches. We use the ode45 function in MATLAB to solve the dynamical approaches in all our numerical experiments.

5.1. In the smooth case.

Example 5.1. Logistic regression [4]: Consider the problem (1.1) as follows:

\[
\begin{align*}
\min f(x) &= \log \left( 1 + \exp \left( - (1, 1, 1)^T x \right) \right) \\
\text{s.t. } Ax &= b, x \in \mathcal{X},
\end{align*}
\]

(5.1)

where \(A = \begin{bmatrix} 0.2, 1, 1, 2 \\ 0, 1, 0.5, 1 \end{bmatrix}, \ b = \text{col} (1, 1), \ \mathcal{X} = \left\{ x \in \mathbb{R}^4_+ | \sum_{i=1}^4 x_i = 1 \right\}. \) The objective function is convex (but not strongly convex) and smooth, and it is a very popular regularization in machine learning. Applying APDMD (4.1) with Kullback-Leibler divergence to address problem (5.1). Figure 4 (left) shows the trajectories of estimates for \(x(t)\) of APDMD (4.1) versus time; Figure 4 (middle) and (right) display the errors of \(|f(x(t)) - f(x^*)|\) and \(\|Ax(t) - b\|\) respectively. The numerical results of gaps for objective function and equation constraints are in excellent agreement with our theoretical results, i.e., they both converge at the predicted rates.

Example 5.2. Distributed Logistic regression: Consider a distributed problem (3.1) as follows:

\[
\begin{align*}
\min f(x) &= \sum_{i=1}^n \log \left( 1 + \exp \left( - \left( i - 1, i, i + 1 \right)^T x_i \right) \right) \\
\text{s.t. } Lx &= 0, x \in \prod_{i=1}^n \mathcal{X}_i,
\end{align*}
\]

(5.2)

where \(x = (x_1, ..., x_n)^T \in \mathbb{R}^{4n}\). Setting \(n = 4\) and \(\mathcal{X}_1 = \left\{ x_1 \in \mathbb{R}^4_+ | 1^T x = 1 \right\}, \ \mathcal{X}_2 = \left\{ x_2 \in \mathbb{R}^4_+ \right\}, \ \mathcal{X}_3 = \left\{ x_3 \in \mathbb{R}^4 \left| x_3 - (0.1, 0.2, 0.5, 0.8)^T \right| \leq 2 \right\} \) and \(\mathcal{X}_4 = \left\{ x_4 \in \mathbb{R}^4 \left| (1, 1, 1, 1) x_4 \leq 4 \right. \right\}. \) Note that the objective function in (5.2) is convex (but not strongly convex) and smooth, which satisfies the requirement of problem (3.1) in the smooth case. Applying ADPDMD (3.20) with 4 agents (connected as a ring) to solve problem (5.2), i.e., for agent 1, using Kullback-Leibler divergence; for agent 2, applying Itakura-Saito divergence; for agent 3, adopting projection operator of Sphere set; for agent 4, utilizing projection operator of half-space set. Figure 5 (left) shows the trajectories of \(x(t)\) under ADPDMD (3.20) for problem (5.2) are uniformly convergent and globally asymptotically stable, i.e., \(x_1 = x_2 = ... = x_5\); Figure 5 (middle) and (right) display the errors of \(|f(x(t)) - f(x^*)|\) and \(\sqrt{x(t)^T Lx(t)}\), respectively. The
Example 5.3. **Distributed quadratic programming**: Consider a special case of DEMO (3.2) in the smooth case as follows:

\[
\min f(x) = \sum_{i=1}^{n} x_i^T A_i x_i,
\]

(5.3)

s.t. \( x + Ly = 0, x \in \prod_{i=1}^{n} \mathcal{X}_i \),

where \( \mathcal{X}_i = \{ x_i \in \mathbb{R}^m | i + 2 \leq x_{i,j} \leq i + 3, j = 1, \ldots, m \} \), \( i = 1, \ldots, n \), \( b = \text{col} (7, \ldots, 7) \in \mathbb{R}^{nm} \) and \( y \in \mathbb{R}^{nm} \) is an auxiliary variable. Let \( n = 10 \), \( m = 5 \) and \( A_i \in \mathbb{R}^{10 \times 10} \) be a positive semi-definite matrix generated by standard Gaussian distribution. Let every part \( x_i \) be an agent, and 10 agents are connected as a ring. Applying the ADMD (3.25) to solve problem (5.3) and the experimental results are shown in Figure 6. In Figure 6 (left), the dual variable \( \lambda \) is uniformly convergent and globally asymptotically stable, i.e., \( \lambda_1 = \lambda_2 = \ldots = \lambda_{10} \); Figure 6 (middle and right) illustrate \( |f(x^\mu(t)) - f(x^*)| \)
and $\sqrt{\lambda(t)^T L\lambda(t)}$ of ADMD (3.25) are convergent at the predicted rates.

5.2. In the nonsmooth case.

Example 5.4. Nonnegative Basis Pursuit (NBP): [34] Take into account a NBP as follows:

\[
\begin{align*}
\min f(x) &= \|x\|_1, \\
\text{s.t. } &Ax = b, x \in \mathcal{X},
\end{align*}
\]

where $\mathcal{X} = \{x \in \mathbb{R}^{40}, x_i \geq 0, i = 1, \ldots, 40\}$ and $A \in \mathbb{R}^{10 \times 40}$ is an orthogonal Gaussian matrix. The objective function in (5.4) is convex (but not strongly convex) and nonsmooth as required. Applying the SAPDMD (4.1) and let $\psi(x) = \sum_{i=1}^{n} x_i \ln x_i$ to solve problem (5.4). Figure 7 displays the reconstructed sparse signal of $x$ in the left, the error of $\|f(x^\mu(t)) - f(x^*)\|$ in the middle, and the error of $\|Ax^\mu(t) - b\|$ in the right. All parameters are complied with the requirement. The numerical results are in good accordance with our theoretical results, where the $\|f(x^\mu(t)) - f(x^*)\|$ and $\|Ax^\mu(t) - b\|$ are convergent at the predicted rates.

Example 5.5. Distributed basis pursuit in row partition [55]: Take a dis-
ACCELERATED PRIMAL-DUAL MIRROR

Fig. 8. SADPDMD (4.12) with with $\alpha = 3$. (left) Trajectories of $x$ in 5 agents; (middle) Reconstructed sparse signal of 5 agents; (right) Error of $|f(x^*(t)) - f(x^*)|$. 

Fig. 9. SADMD (4.16) with with $\alpha = 3$. (left) Trajectories of $\lambda$ of 10 agents; (middle) Reconstructed sparse signal; (right) Error of $|f(x^*(t)) - f(x^*)|$. 

tributed basis pursuit with row partition of sensing matrix as follows:

$$
\min f(x) = \sum_{i=1}^{k} \|x_i\|_1
$$

s.t. $Lx = 0, x \in \mathcal{X},$

where $L = L_k \otimes I_n \in \mathbb{R}^{kn \times kn}$, $\mathcal{X} = \{x \in \mathbb{R}^{kn} | \hat{A}x = b, \hat{A} = \text{bldiag} \{A_{m_1 \times n}, ..., A_{m_k \times n}\} \in \mathbb{R}^{m \times kn}, b \in \mathbb{R}^m\}, b \in \mathbb{R}^m$. Note that, the problem (5.5) is convex (but not strongly convex) and nonsmooth and as a special case of DCCP (3.1) in the nonsmooth case. Setting $m = 10, k = 5, n = 60$, sparsity be 2 and every $A_{m_i \times n}, i = 1, ..., 5$ acts as an agent, and 5 agents are connected as a ring. Applying the SADPDMD (4.12) with $\nabla \psi^* = P_{\mathcal{X}}$ of affine set $\mathcal{X} = \{x \in \mathbb{R}^{kn} | \hat{A}x = b \}$ in Subsection 2.4 to solve the problem (5.5) and all parameters comply with the requirement. The numerical results are displayed in Figure 8. As can be seen from Figure 8 (left) that the variable $x$ is uniformly convergent and globally asymptotically stable, i.e., $x_1^\mu = x_2^\mu = ... = x_5^\mu$; Figure 7 (middle) shows that the SADPDMD can efficiently solve the problem (5.5) and reconstruct the original sparse signal; the result in Figure 8 (middle) is in good accordance with our theoretical results, i.e., $|f(x^*(t)) - f(x^*)|$ is convergent at the predicted rates.
Example 5.6. Distributed basis pursuit in column partition [55]: Consider a distributed basis pursuit with column partition of sensing matrix as follows:

\[
\min f(x) = \sum_{i=1}^{q} \|x_i\|_1,
\]

s.t. \( \bar{A}x + Ly = b \),

where \( \bar{A} = \text{bldiag}\{A_{m \times n_1}, \ldots, A_{m \times n_q}\} \in \mathbb{R}^{mq \times n}, \sum_{i=1}^{q} n_q = n \), and \( y \in \mathbb{R}^{mq} \) is an auxiliary variable. Note that, the problem (5.6) is convex (but not strongly convex) and nonsmooth, which is a special case of DEMO (3.2). Letting \( q = 10 \), and \( A \in \mathbb{R}^{10 \times 60} \), then \( \hat{A} \in \mathbb{R}^{100 \times 60} \). Let every part \( A_{m \times n_i}, i = 1, \ldots, 10 \) be an agent, and 10 agents are connected as a ring. Applying the ADMD (4.16) to solve the problem (5.6) and the experimental results are shown in Figure 9. Figure 9 (left) displays the dual variable \( \lambda \) is uniformly convergent and globally asymptotically stable, i.e., \( \lambda_1^\mu = \lambda_2^\mu = \ldots = \lambda_{10}^\mu \); Figure 9 (middle) describes that the SADMD (4.16) is able to efficiently solve the problem (5.6) and recover the sparse signal in a distributed way; Figure 9 (right) illustrates \( |f(x^\mu(t)) - f(x^*)| \) of SADMD (4.16) converges at the predicted rates.

6. Conclusions. In this paper, we have proposed two accelerated primal-dual mirror dynamical approaches to the problem (1.1) in the smooth and nonsmooth cases. For the problem (1.1) in the smooth case, we first proposed an APDMD on account of mirror dynamical descent and primal-dual framework, then, by the Lyapunov analysis method, we studied the accelerated convergence rate of the trajectories by the APDMD. We have extended the APDMD to two distributed networks optimization problems (DCCP and DEMO) and obtained accelerated convergence rates of them. Then, for the problem (1.1) in the nonsmooth case, we have studied a SAPDMD (smoothing version of APDMD) and investigated the convergence rates of SAPDMD under sophisticated smoothing approximation parameters. Finally, we have also established the convergence result of SAPDMD when applying it to deal with DCCP and DEMO. The obtained results can be straightforwardly transferred to inertial (primal-dual) dynamical approaches for the problems (1.3), (1.5) and (1.7).

REFERENCES

[1] F. Alimisis, A. Orvieto, G. B’ecigneul, and A. Lucchi, A continuous-time perspective for modeling acceleration in riemannian optimization, in AISTATS, 2020.
[2] F. Alvarez, On the minimizing property of a second order dissipative system in hilbert spaces, SIAM Journal on Control and Optimization, 38 (2000), pp. 1102–1119.
[3] V. Apidopoulos, J.-F. Aujol, and C. Dossal, The differential inclusion modeling fista algorithm and optimality of convergence rate in the case \( b \leq 3 \), SIAM Journal on Optimization, 28 (2018), pp. 551–574.
[4] H. Attouch, Z. Chbani, J. M. Fadili, and H. Riahi, Fast convergence of dynamical admm via time scaling of damped inertial dynamics, Journal of Optimization Theory and Applications, (2021).
[5] H. Attouch, Z. Chbani, and H. Riahi, Rate of convergence of the Nesterov accelerated gradient method in the subcritical case \( \alpha \leq 3 \), ESAIM: Control, Optimisation and Calculus of Variations, 25 (2019), p. 2.
[6] J. Aujol, C. Dossal, and A. Rondepierre, Convergence rates of the Heavy-ball method under the Lojasiewicz property, Mathematical Programming, (2022), pp. 1–60.
[7] J.-F. Aujol, C. Dossal, and A. Rondepierre, Convergence rates of the Heavy-ball method for quasi-strongly convex optimization, (2020), https://hal.archives-ouvertes.fr/hal-02545245v2.
[8] A. Balavoine, C. J. Rozell, and J. K. Romberg, Convergence speed of a dynamical system for sparse recovery, IEEE Transactions on Signal Processing, 61 (2013), pp. 4259–4269.
[9] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, Clustering with Bregman divergences, Journal of Machine Learning Research, 6 (2005), pp. 1705–1749.
[10] H. H. Bauschke and P. L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces, in CMS Books in Mathematics, 2011.
[11] A. Beck and M. Teboulle, Mirror descent and nonlinear projected subgradient methods for convex optimization, Operations Research Letters, 31 (2003), pp. 167–175.
[12] P. Bégout, J. Bolte, and M. A. Jendoubi, On damped second-order gradient systems, Journal of Differential Equations, 259 (2015), pp. 3115–3143.
[13] W. Bian and X. Chen, Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation, IEEE Transactions on Neural Networks and Learning Systems, 25 (2014), pp. 545–556.
[14] W. Bian and X. Chen, A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty, SIAM Journal on Numerical Analysis, 58 (2020), pp. 858–883.
[15] J. Bolte, Sur des systèmes dynamiques dissipatifs de type gradient. applications en optimisation, Thèse, Université Montpellier 2.
[16] S. P. Boyd, N. Parikh, E. K. Wah Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via alternating direction method of multipliers, Foundations and Trends in Machine Learning, 3 (2011), pp. 1–122.
[17] R. I. Bot and E. R. Csetnek, Second order forward-backward dynamical systems for monotone inclusion problems, SIAM Journal on Control and Optimization, 54 (2016), pp. 1423–1443.
[18] R. I. Bot and D.-K. Nguyen, Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping, Journal of Differential Equations, (2021).
[19] A. Cabot and P. Frankel, Asymptotics for some semilinear hyperbolic equations with nonautonomous damping, Journal of Differential Equations, 252 (2012), pp. 294–322.
[20] A. Cabot and L. Paoli, Asymptotics for some vibro-impact problems with a linear dissipation term, Journal de mathématiques pures et appliquées, 7 (2007), pp. 2030037–2030038.
[21] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, Mathematical Programming, 134 (2012), pp. 71–99.
[22] J. Diakonikolas and M. I. Jordan, Generalized momentum-based methods: A Hamiltonian perspective, SIAM Journal on Optimization, 31 (2021), pp. 915–944.
[23] M. Elouami, R. May, and C. Mesnou, Asymptotic for a second order evolution equation with vanishing damping term and Tikhonov regularization, arXiv:1711.11241, 2017.
[24] M. Fazlyab, A. Koppel, V. M. Preciado, and A. Ribeiro, A variational approach to dual methods for constrained convex optimization, 2017 American Control Conference (ACC), (2017), pp. 5269–5275.
[25] D. Fellier and F. Paganini, Stability of primal–dual gradient dynamics and applications to network optimization, Automatica, 46 (2010), pp. 1974–1981.
[26] C. D. Godsil and G. F. Royle, Algebraic graph theory, Graduate texts in mathematics, 2001.
[27] X. He, D. W. Ho, T. Huang, J. Yu, H. Abu-Rub, and C. Li, Second-order continuous-time algorithms for economic power dispatch in smart grids, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48 (2017), pp. 1482–1492.
[28] X. He, T. Huang, J. Yu, C. Li, and C. Li, An inertial projection neural network for solving variational inequalities, IEEE Transactions on Cybernetics, 47 (2016), pp. 809–814.
[29] X. He, Y. Zhao, and T. Huang, Optimizing the dynamic economic dispatch problem by the distributed consensus-based ADMM approach, IEEE Transactions on Industrial Informatics, 16 (2020), pp. 3210–3221.
[30] X.-F. He, R. Hu, and Y. Fang, Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems, SIAM Journal on Control and Optimization, 59 (2021), pp. 3278–3301.
[31] X.-F. He, R. Hu, and Y. Fang, Fast convergence of primal-dual dynamics and algorithms with time scaling for linear equality constrained convex optimization problems, arXiv:2103.12931, 2021.
[32] J. J. Hopfield and D. W. Tank, Computing with neural circuits: a model, Science, 233 4764 (1986), pp. 625–33.
[33] M. P. Kennedy and L. O. Chua, Neural networks for nonlinear programming, IEEE Transactions on Circuits and Systems, 35 (1988), pp. 554–562.
[34] M. A. Kriafehnejad, A. G. Dimakis, W. Xu, and B. Hassibi, Sparse recovery of nonnegative signals with minimal expansion, IEEE Transactions on Signal Processing, 59 (2010),
W. Krichene, A. Bayen, and P. Bartlett, *A Lyapunov Approach to accelerated first-order optimization in continuous and discrete time*, PhD thesis, University of California, Berkeley, 2016.

W. Krichene, A. M. Bayen, and P. L. Bartlett, *Accelerated mirror descent in continuous and discrete time*, in Conference on Neural Information Processing Systems, 2015.

Z. Lin, H. Li, and C. Fang, *Accelerated optimization for machine learning*, Springer, 2020.

Q. Liu, S. Yang, and J. Wang, *A collective neurodynamic approach to distributed constrained optimization*, IEEE Transactions on Neural Networks and Learning Systems, 28 (2017), pp. 1747–1758.

H. Luo, *Accelerated primal-dual methods for linearly constrained convex optimization problems*, arXiv:2109.12604, 2021.

W. Michiels, T. Vyhlidal, H. Hulberts, and H. Nijmeijer, *Stabilizability and stability robustness of state derivative feedback controllers*, SIAM Journal on Control and Optimization, 47 (2009), pp. 3100–3117.

Y. Nesterov, *Smooth minimization of non-smooth functions*, Mathematical Programming, 103 (2005), pp. 127–152.

D. O’Connor and L. Vandenberghe, *Primal-dual decomposition by operator splitting and applications to image deblurring*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 1724–1754.

N. Parikh and S. P. Boyd, *Proximal algorithms*, Foundations and Trends in Optimization, 1 (2014), pp. 127–239.

B. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Computational Mathematics and Mathematical Physics, 4 (1964), pp. 1–17.

G. Qu and N. Li, *On the exponential stability of primal-dual gradient dynamics*, IEEE Control Systems Letters, 3 (2019), pp. 43–48.

R. T. Rockafellar, *Convex analysis*, in Princeton Landmarks in Mathematics and Physics, 1970.

W. J. Su, S. P. Boyd, and E. J. Candès, *A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights*, Journal of Machine Learning Research, 17 (2016), pp. 1–43.

A. Wibisono, A. C. Wilson, and M. I. Jordan, *A variational perspective on accelerated methods in optimization*, Proceedings of the National Academy of Sciences, 113 (2016), pp. E7351–E7358.

P. Yi, Y. Hong, and F. Liu, *Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems*, Automatica, 74 (2016), pp. 259–269.

X. Zeng, J. Lei, and J. Chen, *Dynamical primal-dual accelerated method with applications to network optimization*, arXiv:1912.03690, 2019.

X. Zeng, P. Yi, Y. Hong, and L. Xie, *Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems*, SIAM Journal on Control and Optimization, 56 (2018), pp. 3973–3993.

Y. Zhao, X. He, T. Huang, and Q. Han, *Analog circuits for solving a class of variational inequality problems*, Neurocomputing, 295 (2018), pp. 142–152.

Y. Zhao, X. He, T. Huang, and J. Huang, *Smoothing inertial projection neural network for minimization Lp-q in sparse signal reconstruction*, Neural networks, 99 (2018), pp. 31–41.

Y. Zhao, X. Liao, X. He, R. Tang, and W. Deng, *Smoothing inertial neurodynamic approach for sparse signal reconstruction via Lp-norm minimization*, Neural Networks, 140 (2021), pp. 100–112.

Y. Zhao, X.-F. Liao, X. He, and R. Tang, *Centralized and collective neurodynamic optimization approaches for sparse signal reconstruction via L1-minimization*, IEEE Transactions on Neural Networks and Learning Systems, 2021, http://dx.doi.org/10.1109/TNNLS.2021.3085314.

Z. A. Zhu and E. Hazan, *Optimal black-box reductions between optimization objectives*, in Conference on Neural Information Processing Systems, 2016.