Journal of Machine Learning Research 24 (2023) 1-59 Submitted 9/22; Revised 8/23; Published 11/23

Accelerated Primal-Dual Mirror Dynamics for Centralized and Distributed Constrained Convex Optimization Problems

You Zhao zhaoyou1991sdtz@163.com
College of Electronic and Information Engineering
Southwest University
Chongqing 400044, PR. China

Xiaofeng Liao xfliao@cqu.edu.cn
College of Computer Science
Chongqing University
Chongqing 400044, PR. China

Xing He hexingdoc@swu.edu.cn
College of Electronic and Information Engineering
Southwest University
Chongqing 400715, PR. China

Mingliang Zhou mingliangzhou@cqu.edu.cn
College of Computer Science
Chongqing University
Chongqing 400044, PR. China

Chaojie Li cjlee.cqu@163.com
State Key Laboratory of Alternate Electrical Power System With Renewable Energy Sources
North China Electric Power University
Beijing 102206, China

Editor: Martin Jaggi

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 along with trajectories 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 and comparative experiments are given to demonstrate the effectiveness and superiority of the proposed accelerated mirror dynamical approaches.

Keywords: machine learning, accelerated mirror dynamical approaches, constrained smooth and nonsmooth convex optimization, smoothing approximation, distributed approaches

1. Introduction

1.1 Problem statement

In this paper, we consider the following convex constrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x), \text{ s.t. } Ax = b, x\in\mathcal{X}, \tag{1}$$

where

$$\begin{cases} \mathcal{X} \text{ is a closed and convex set,} \\ f:\mathcal{X}\to\mathbb{R} \text{ may be a smooth or nonsmooth convex function,} \\ A:\mathcal{X}\to\mathbb{R}^m \text{ is a continuous linear operator and } b\in\mathbb{R}^m, \\ \text{The optimal solution set of problem (1) is nonempty.} \end{cases} \tag{2}$$

This problem covers many optimization problems in various applied fields such as machine learning, sparse signal reconstruction, image deblurring, resource allocation, saddle-point problems, Markov decision processes, regularized empirical risk minimization, and supervised machine learning. (see, e.g. Boyd et al. (2011), Lin et al. (2020),Darbon and Langlois (2021), O'Connor and Vandenberghe (2014), Nandwani et al. (2019), Zhang and Xiao (2017), Tiapkin and Gasnikov (2022), Li and Lin (2020)).

1.2 Historical presentation

The inertial (second-order or accelerated) dynamical approaches are increasingly popular for solving the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x). \tag{3}$$

To deal with problem (3), Polyak (1964) first proposed the heavy ball with friction (HBF) dynamical approach

$$\text{(HBF)} \quad \ddot{x}(t) + \eta\dot{x}(t) + \nabla f(x) = 0, \tag{4}$$

where $\eta > 0$ is a damping parameter. Later, Alvarez (2000) studied the asymptotic behavior of (4) with a time independent parameter η when $f(x)$ is convex. Bégout et al. (2015) studied some convergence properties of the HBF (4) with a constant η when $f(x)$ is a nonconvex function. Aujol et al. (2022, 2023) investigated the convergence rate of HBF dynamical system and its corresponding discrete algorithm with a fixed parameter η under the condition that $f(x)$ satisfies quasi-strongly convex and Lojasiewicz properties, respectively. When we replace the constant η with the function $\frac{\alpha}{t}$ with $\alpha\geq 3$ in HBF (4), Su et al. (2016) first revealed that it can be regarded as the continuous-time limit of the Nesterov's accelerated gradient algorithm and it has an accelerated convergence rate, i.e,

$f(x(t)) - \min f(x) = O\left(\frac{1}{t^2}\right)$. When $\alpha > 3$, the weak convergence of trajectories $x(t)$ to minimizers of f, and improved convergence rate $f(x(t)) - \min f(x) = o\left(\frac{1}{t^2}\right)$ have been proved in May (2017) and Attouch et al. (2018a). By introducing a Tikhonov regularization, improved convergence rate $f(x(t)) - \min f(x) = o\left(\frac{1}{t^2}\right)$ and the strong convergence of the trajectories to the element of minimum norm of the set of minimizers of f have been shown by Elloumi et al. (2017) and Attouch et al. (2018b) under suitable conditions. When $\eta = \frac{\alpha}{t}, 0 < \alpha \leq 3$, the convergence rate $f(x(t)) - \min f(x) = O\left(t^{-\frac{2\alpha}{3}}\right)$ was estimated by Attouch et al. (2019) and Vassilis et al. (2018) for smooth and nonsmooth convex functions $f(x)$. Cabot and Frankel (2012) studied the asymptotic behavior of the HBF (4) when replacing η by $\frac{\alpha}{t^\gamma}, 0 < \gamma < 1$. In addition. May et al. (2021) further investigated long time behavior of the trajectories $x(t)$ of a second order evolution equation with damping and regularizing terms, which contains HBF (4) when replacing η by $\frac{\alpha}{t^\gamma}, 0 < \gamma < 1$ and regularizing terms. Moreover, Wibisono et al. (2016) studied a series of accelerated (inertial) dynamical approaches to the problem (3) by the Bregman Lagrangian based on the calculus of variations. Kovachki and Stuart (2021) studied the behavior of momentum methods for solving the problem (3) from a continuous-time perspective. Wilson et al. (2021) proposed several Lyapunov functions for analyzing the accelerated (momentum) algorithms to solve the problem (3). By the manifold with curvature bounded from below, Alimisis et al. (2020) proposed a Riemannian variant of accelerated gradient dynamical approach.

Recently, to solve the problem (3) with a closed and convex set constraint $\mathcal{X}$, i.e.,

$$\min_{x\in\mathbb{R}^n} f(x), \text{ s.t. } x \in \mathcal{X} \subset \mathbb{R}^n, \tag{5}$$

many inertial dynamical approaches have been studied. Based on the projection operators, He et al. (2016) proposed an inertial dynamical approach to solve the problem (5) with non-convex objective functions. Later, based on the work in (He et al. (2016)) and smoothing approximation technique, a smoothing inertial projection neurodynamic approach was proposed to solve constrained non-convex L_{p-q} minimization problem to reconstruct the sparse signal in (Zhao et al. (2018)). Combining the dynamical approach of mirror descent with the continuous version of Nesterov's acceleration algorithm, Krichene et al. (2015) proposed an accelerated mirror dynamical approach for solving the problem (5) as follows:

$$\begin{cases} \dot{X} = \frac{\gamma}{t}\left(\nabla\psi^*(Z) - X\right), \\ \dot{Z} = -\frac{t}{\gamma}\nabla f(X), \\ X(0) = x_0 = \nabla\psi^*(z_0), Z(0) = z_0, \end{cases} \tag{6}$$

where $\gamma \geq 2$ and $\nabla\psi^*$ is the gradient of conjugate of ψ (see **Section 2.3**).

In addition, in order to solve the problem (3) with an affine constraint, i.e,

$$\min_{x\in\mathbb{R}^n} f(x), \text{ s.t. } Ax = b, \tag{7}$$

many inertial dynamical approaches under primal-dual framework were extensively investigated. Zeng et al. (2022) first proposed a second-order dynamical primal-dual approach for solving the problem (7) and proved that the convergence rate of the gap of the Lagrangian function is $O\left(\frac{1}{t^2}\right)$. The corresponding inertial dynamical approach is then extended to solve two distributed optimization problems. Later, Boţ and Nguyen (2021) improved the convergence rates of the works in (Zeng et al. (2022)) and provided the weakly convergence analysis to a primal-dual optimal solution of the problem (7). He et al. (2021) studied an inertial primal-dual dynamical approach without/with perturbations for separable convex optimization problems with an affine constraint. Attouch et al. (2022) 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, Luo (2021) further proposed a “second-order”+“first-order” primal-dual dynamical approaches for solving problem (7) with accelerated convergence properties. Recently, when the f is strongly convex in the problem (7), many accelerated or optimal primal-dual discrete-time (numerical) algorithms have been proposed. Salim et al. (2022) proposed an accelerated primal-dual discrete-time algorithm and provided lower bounds for gradient computation and matrix multiplication, and proved that the proposed algorithm is the first optimal algorithm for this class of problems. An accelerated primal-dual algorithm for solving smooth convex-concave saddle-point problems with a structure has been proposed in (Alkousa et al. (2020)). Kovalev et al. (2022) proposed an accelerated primal-dual gradient method (APDG) for solving smoothing convex-concave saddle-point problems with bilinear coupling, and obtained the accelerated or optimal linear convergence rates when the objective functions satisfy certain conditions. Sadiev et al. (2022) proposed an accelerated primal-dual algorithm with inexact prox for communication acceleration, and provided a new convergence rate for strongly convex-concave saddle-point problems with bilinear coupling characterized without smoothness in the dual function. Kovalev et al. (2020) investigated two new decentralized discrete-time algorithms for decentralized optimization problem with smooth strongly convex objective functions based on primal-dual algorithms, where the first algorithm is optimal both in terms of the number of communication rounds and in terms of the number of gradient computations, and the second algorithm is optimal terms in terms of communication rounds.

This paper aims to investigate accelerated primal-dual dynamical approaches based on mirror descent and smoothing approximation methods for solving problem (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) 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) with accelerated convergence rates of primal-dual, objective functions gaps and feasibility measure. We provide the 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 constraints), existence and uniqueness of the global solution of APDMD are obtained. Last, applying the APDMD to address the DCCP and EDMO leads to two distributed APDMDs (i.e., ADPDMD and ADMD) with accelerated convergence properties.

- For the problem (1) in the nonsmooth case, a smoothing accelerated primal-dual mirror dynamical approach (SAPDMD) for the problem (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 conditions. Last, applying the SAPDMD to DCCP and EDMO leads to two distributed SAPDMDs (i.e., SADPDMD and SADMD) with accelerated convergence gurantee.

- Last but not least, the asymptotic analysis and the obtained results in this paper can be straightly forwardly transferred to inertial dynamical approaches for the problems (3), (5) and (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) 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) 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 gives some essential mathematical preliminaries.

2.1 Convex analysis

A function $f : \mathcal{X} \to \mathbb{R}$ is convex, if it satisfies $\theta f(x) + (1-\theta) f(z) \geq f(\theta x + (1-\theta) z), \forall x, z \in \mathcal{X}, x \neq z$, and $\theta \in (0,1)$. The subdifferential $g_f(x)$ of $f(x)$ with respect to $x \in \mathcal{X}$ is defined by $g_f(x) = \left\{ p \in \mathbb{R}^n | f(z) - f(x) \geq p^T (z-x), \forall x, z \in \mathcal{X} \right\}$, 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 $\mathcal{L}_\beta : \mathcal{X} \times \mathbb{R}^m \to \mathbb{R}$ with respect to the problem (1) with $\beta \geq 0$ is defined by

$$\mathcal{L}_\beta(x, \lambda) = f(x) + \lambda^T (Ax - b) + \frac{\beta}{2} \|Ax - b\|^2, \tag{8}$$

Then, (x^*, λ^*) is an optimal solution of the problem (1) if and only if it is a saddle point of $\mathcal{L}_\beta$, i.e.,

$$\mathcal{L}_\beta(x^*, \lambda) \leq \mathcal{L}_\beta(x^*, \lambda^*) \leq \mathcal{L}_\beta(x, \lambda^*), \forall (x, \lambda) \in \mathcal{X} \times \mathbb{R}^m. \tag{9}$$

2.3 Dual distance and Bregman divergence

Define $\psi : \mathcal{X} \to \mathbb{R}$ and let $\mathcal{X}$ be a closed and convex set , then, by the *Legendre-Fenchel transform*, the conjugate function of ψ is

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

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

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

from the Fenchel's duality theorem.

Assume that ψ and ψ^* 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 (see Rockafellar (1997)). In addition, since $\psi^{**} (u) = \psi (x)$, one has

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

which implies $\partial\psi^* (u) = arg\max_{x \in \mathcal{X}} \left\{ u^T x - \psi (x) \right\}$. Since $\mathrm{dom}\psi = \mathcal{X}$, then $\partial\psi^* (u) \subset \mathcal{X}$, i.e., the set-valued map $\partial\psi^*$ can map $\mathcal{X}^*$ into $\mathcal{X}$. In order to make $\partial\psi^* (u) \subset \mathcal{X}$ be unique, i.e., the $\psi^* (u)$ is differentiable for any $u \in \mathcal{X}^*$, the following definitions are needed (see Krichene et al. (2016)).

Definition 1. *A convex function ψ is cofinite if its epigraph does not consist of any non-vertical half-line.*

Definition 2. *A convex function ψ is essentially strictly convex if it is strictly convex and subdifferentiable on all convex subsets.*

Lemma 3. *If ψ and ψ^* are proper, convex, and closed, so they are inverses of each other, then ψ^* is finite and differentiable on $\mathcal{X}^*$ if and only if ψ is essentially strictly convex and cofinite.*

Lemma 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 when z closes y, i.e., $D_h (y, z) = \frac{1}{2} (y - z)^T \nabla^2 h (z) (y - z) + o\left(\|y - z\|^2\right) := \frac{1}{2} \|y - z\|_{\nabla^2 h(z)}, \forall y, z \in \mathcal{X}$ (see Wibisono et al. (2016)).

2.4 Projection operators

Define the projection operator of a closed and convex set $\mathcal{X}$ at x be $P_{\mathcal{X}} (u) = \arg\min_{z \in \mathcal{X}} \|u - z\|$. A basic property of the projection operator $P_{\mathcal{X}}$ is

$$(P_{\mathcal{X}} (x) - z)^T (P_{\mathcal{X}} (x) - x) \leq 0, \forall x \in \mathbb{R}^n, z \in \mathcal{X}.$$

Lemma 5. *1): If $\mathcal{X}$ is a **box** set, i.e. $\mathcal{X} = \{u \in \mathbb{R}^n \mid \underline{u}_i \leq u_i \leq \bar{u}_i,\ i = 1, ..., n\}$, then, for the i-th component of $P_{\mathcal{X}}(u)$, we have*

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

*2): If $\mathcal{X}$ is a **ball** 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 $P_{\mathcal{X}}(u) = \begin{cases} u, & \text{if } \|u - z\| \leq r, \\ z + \frac{r(u-z)}{\|u-z\|}, & \text{if } \|u - z\| > r, \end{cases}$ where $\|\cdot\|$ denotes the 2-norm, i.e., $\|u\| = \sqrt{\sum_{i=1}^n u_i^2}$ in the whole paper without any special note.*

*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* $\mathrm{Rank}(A) < m$, *and $A^{\dagger} = A^T \left(AA^T\right)^{-1}$ when* $\mathrm{Rank}(A) = m$.

*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 $P_{\mathcal{X}}(u) = \begin{cases} u + \frac{b - a^T u}{\|a\|^2} a, & \text{if } a^T u > b, \\ u, & \text{if } a^T u \leq b. \end{cases}$ For more information regarding the projection operators, please refer to (Bauschke and Combettes; Parikh et al. (2014)).*

2.5 Graph Theory

An undirected communication topology graph is a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with node set $\mathcal{V} = \{\nu_1, \nu_2, ..., \nu_n\}$, edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ and connection matrix $\mathcal{A} = \{\mathrm{a}_{ij}\}_{n\times n}$ with nonnegative elements $\mathrm{a}_{ij} = \mathrm{a}_{ji} > 0$ if $(i, j) \in \mathcal{E}$, and $\mathrm{a}_{ij} = \mathrm{a}_{ji} = 0$ 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, $\cdots$, 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 $\mathcal{G}$ is connected if there exists a path between any pair of distinct nodes v_i and v_j $(i, j = 1, 2, ..., n)$ (see Godsil and Royle (2001)).

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 6. *Let $\hat{f} : \mathcal{X} \times [0, \bar{\mu}] \to \mathbb{R}$ with $\bar{\mu} > 0$ be a smoothing function of the convex function $f : \mathcal{X} \to \mathbb{R}$, if $\hat{f}(\cdot, \mu)$ is continuous differentiable for any fixed $\bar{\mu} \geq \mu > 0$, then it enjoys the following properties (see Bian and Chen (2020); Chen (2012))*

(i) (approximation property) $\lim\limits_{x\to w, \mu\to 0} \hat{f}(x, \mu) = f(w), \forall\ w \in \mathcal{X}$;

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

(iii) (gradient consistency) $\left\{\lim\limits_{x\to w, \mu\to 0} \nabla_x \hat{f}(x, \mu)\right\} \subseteq \partial f(w), \forall\ w \in \mathcal{X}$;

(iv) (gradient boundedness and Lipschitz continuous of μ) There exists a positive constant $\kappa_{\hat{f}} > 0$ such that

$$\left|\nabla_\mu \hat{f}(x,\mu)\right| \le \kappa_{\hat{f}} \quad \forall\ \mu \in [0,\bar{\mu}], \forall\ x \in \mathcal{X},$$

and for any $x \in \mathcal{X}$,$u_1, \mu_2 \in [0,\bar{\mu}]$, it follows

$$\left|\hat{f}(x,\mu_1) - \hat{f}(x,\mu_2)\right| \le \kappa_{\hat{f}} |\mu_1 - \mu_2|;$$

(v) (Lipschitz continuity with respect to x) there is a constant ℓ such that for any fixed $\mu \in (0,\bar{\mu}]$, $\nabla_x \hat{f}(x,\cdot)$ (i.e., the gradient of $\hat{f}(x,\mu)$ with respect to x when μ is fixed) is Lipschitz continuous with respect to x on $\mathcal{X}$ with a Lipschitz constant $\frac{\ell}{\mu}$.

In addition, the **Definition 6** *(iv)* implies that $\left|\hat{f}(x,\mu) - f(x)\right| \le \kappa_{\hat{f}}\mu$, $\forall\ 0 < \mu \le \bar{\mu}$, $x \in \mathcal{X}$.

The smoothing function satisfying the above conditions in **Definition 6** enjoys the following properties (see Bian and Chen (2013)):

Lemma 7. *1): If $\hat{f}_1, \hat{f}_2, ..., \hat{f}_n$ are smoothing functions of $f_1, f_2, ..., f_n$, then $\sum_{i=1}^n c_i \hat{f}_i$ is a smoothing function of $\sum_{i=1}^n c_i f_i$ with $\kappa_{\sum_{i=1}^m c_i \hat{f}_i} = \sum_{i=1}^m c_i \kappa_{\hat{f}_i}$ when $c_i \ge 0$ and f_i is regular for any $i = 1, 2, ..., n$.*

2): If $\Phi : \mathcal{X} \to \mathbb{R}$ is locally Lipschitz, $\Psi : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and globally Lipschitz with a Lipschitz constant l_Ψ, then $\Psi\left(\hat{\Phi}\right)$ is a smoothing function of $\Psi(\Phi)$ with $\kappa_{\Psi(\hat{\Phi})} = l_\Psi \kappa_{\hat{\Phi}}$, where $\hat{\Phi}$ is a smoothing function of Φ.

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

Example 1. *The existing results in (Chen (2012)) provide some theoretical basis to construct smoothing functions that satisfy the conditions in **Definition 6**. A smoothing function for the $g(s) = \max\{0, s\}$ is given by*

$$\hat{g}(s,\mu) = \begin{cases} \max\{0,s\}, \text{if } |s| > \mu, \\ \frac{(s+\mu)^2}{4\mu}, \qquad \text{if } |s| \le \mu, \end{cases} \tag{10}$$

with a $\kappa_{\hat{g}} = \frac{1}{4}$. Note that $\hat{g}(s,\mu)$ is convex and nondecreasing of s with any fixed $0 < \mu \le \bar{\mu}$, is also nondecreasing with respect to μ for any fixed s, and $\lim\limits_{\mu\to 0} \hat{g}(s,\mu) = \max\{0,s\}$ (see Figure 1 : (left)).

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

$$\hat{\theta}(s,\mu) = \begin{cases} |s|, \qquad \text{if } |s| > \frac{\mu}{2}, \\ \frac{s^2}{\mu} + \frac{\mu}{4}, \text{if } |s| \le \frac{\mu}{2}, \end{cases} \tag{11}$$

*where $\lim\limits_{\mu\to 0} \hat{\theta}(s,\mu) = |s|$, and $\kappa_{\hat{\theta}} = \frac{1}{4}$ (it is used in **Section 5.2**).*

As can be seen from Figure 1: (right) that $\hat{\theta}(s,\mu)$ is also convex and nondecreasing for any fixed $0 < \mu \le \bar{\mu}$, and nondecreasing with respect of s for any fixed s.

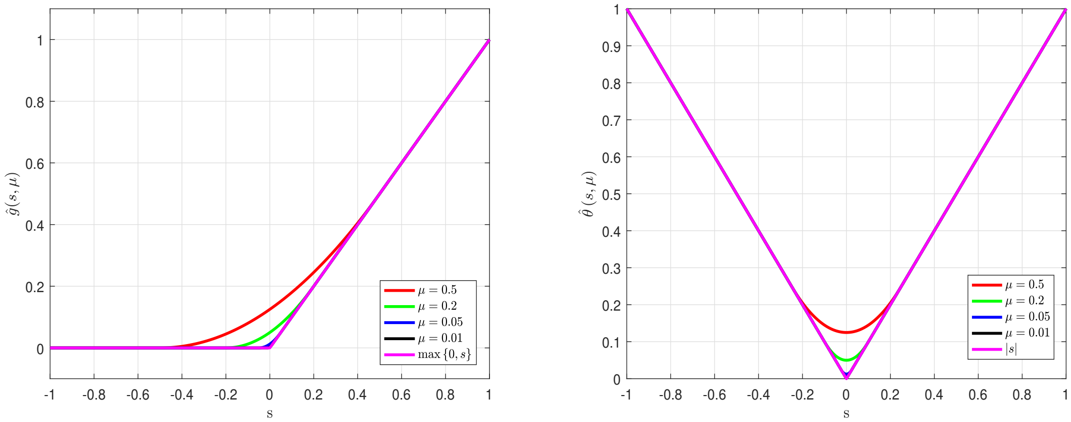

Figure 1: (left) The smoothing function $\hat{g}(s,\mu)$ with different μ. (right) The smoothing function $\hat{\theta}(s,\mu)$ with different μ.

2.7 Notation

Let $\mathbb{R}^n$ ($\mathbb{R}^{m\times n}$) 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, $y \in \mathbb{R}^n$, $x^T y = \sum_{i=1}^{n} x_i y_i$, and the superscript T represents transpose. $\|\cdot\|$ is the Euclidean norm (i.e., 2-norm), $\|\cdot\|_1$ is the 1-norm. $\mathrm{L}^1_{loc}([t_0, +\infty))$ represents the local Lebesgue integral 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 $\bar{A} = \mathrm{bldiag}\{A_1, ..., A_n\} \in \mathbb{R}^{\sum_{i=1}^{n} p_i \times \sum_{i=1}^{n} q_i}$, which means a block diagonal matrix. For $x_i \in \mathbb{R}^m$, $i = 1, ..., n$, we use $\mathrm{col}(x_1, ..., x_n) \in \mathbb{R}^{mn}$ to denote an mn column vector. $\mathbf{1}$ and $\mathbf{0}$ are a scalar or an all-ones column vector and an all-zeros column vector which can be obtained according to the context in the paper.

3. Optimization approaches for problem (1) in the smooth case

In this section, we propose an APDMD approach to address the problem (1) with smooth and convex objective function. Then, we extend it to solve ECCP (12) and DEMO (13) in the smooth case, to obtain ADPDMD (40) and ADMD (50), respectively. To our best knowledge, there are no accelerated mirror dynamical approaches for the problem (1), distributed accelerated dynamical approaches for ECCP (12) and DEMO (13) only with smooth convex objective functions.

To make the well-posedness of the problem (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 ψ is proper, essentially strictly convex and cofinite (see **Section 2.3**);

3. (Slater's condition) There exists a vector $x \in \mathrm{int}(\mathcal{X})$ that satisfies $Ax = b$.

The formulation (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 are to cooperate for seeking the minimum of the following problem:

$$\begin{aligned} &\min_{x\in\mathbb{R}^{nm}} f\left(x\right)=\sum_{i=1}^{n} f_i\left(x_i\right),\\ &\text{s.t. } Lx=0 \ (i.e., x_i=x_j, i,j=1,...,n), x_i\in\mathcal{X}_i\subset\mathbb{R}^m, \end{aligned} \tag{12}$$

where $x=\mathrm{col}\left(x_1,...,x_n\right)\in\mathbb{R}^{nm}$, and $\mathcal{X}=\Pi_{i=1}^n\mathcal{X}_i\subset\mathbb{R}^{nm}$. $L=L_n\otimes I_m\in\mathbb{R}^{nm\times nm}$ 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 $x_i=x_j, i,j=1,2,...,n$ 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 $f_i\left(x_i\right):\mathcal{X}_i\to\mathbb{R}$ and a local feasible constraint set $\mathcal{X}_i\subset\mathbb{R}^{p_i}, i=1,...,n$. Let $x=\mathrm{col}\left(x_1,...,x_n\right)$, $\mathcal{X}=\Pi_{i=1}^n\mathcal{X}_i\subset\mathbb{R}^{\sum_{i=1}^n p_i}$, then, the DEMO problem is

$$\min_{x\in\mathbb{R}^{\sum_{i=1}^n p_i}} f\left(x\right)=\sum_{i=1}^{n} f_i\left(x_i\right),\ \text{s.t.}\ \sum_{i=1}^{n}A_ix_i=\sum_{i=1}^{n}d_i,\ x_i\in\mathcal{X}_i,\ i=1,2,...,n, \tag{13}$$

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

To ensure the well-posedness of DCCP (12) and DEMO (13), some appropriate assumptions need to be made on them, which are fairly standard as follows:

Assumption 3.2

1. The objective function $f\left(x\right)$ is $\sum_{i=1}^m f_i\left(x_i\right)$ and for all $i=1,...,n$, $f_i\left(x_i\right)$ 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 $i=1,...,n$, ψ_i is proper, essentially strictly convex and cofinite (see **Section 2.3**);
3. The communication graph $\mathcal{G}$ is connected and undirected;
4. The Slater's condition of DCCP (12) and DEMO (13) is satisfied.

3.1 APDMD for problem (1) with smooth convex objective functions

Inspired by the accelerated mirror descent in (Krichene et al. (2015)) and primal-dual dynamical approach in (Feijer and Paganini (2010)), we propose the following APDMD:

$$\begin{cases} \dot{x}\left(t\right)=\frac{\alpha}{t}\left(\nabla\psi^*\left(u\left(t\right)\right)-x\left(t\right)\right),\\ \dot{u}\left(t\right)=-\frac{t}{\alpha}\left(\nabla f\left(x\left(t\right)\right)+\beta A^T\left(Ax\left(t\right)-b\right)+A^Tv\left(t\right)\right)-\zeta\dot{x}\left(t\right),\\ \dot{\lambda}\left(t\right)=\frac{\alpha}{t}\left(v\left(t\right)-\lambda\left(t\right)\right),\\ \dot{v}\left(t\right)=\frac{t}{\alpha}\left(A\nabla\psi^*\left(u\left(t\right)\right)-b\right)-\zeta\dot{\lambda}\left(t\right),\\ x\left(t_0\right)=x_0, u\left(t_0\right)=u_0 \text{ with } \nabla\psi^*\left(u_0\right)=x_0\in\mathcal{X}, \lambda\left(t_0\right)=\lambda_0, v\left(t_0\right)=v_0, \end{cases} \tag{14}$$

where $t\geq t_0>0$, $\beta\geq0$, $\zeta\geq0$ and $\alpha\geq2$. The illustration of the APDMD (14) is in Figure 2: (left).

Note that the APDMD (14) can be regarded as the following second-order dynamical approach:

$$
\begin{cases}
\ddot{x}(t) + \frac{\alpha+1}{t}\dot{x}(t) + \nabla^2\psi^*\left(\nabla\psi\left(x(t) + \frac{t}{\alpha}\dot{x}(t)\right)\right) \\
\qquad \times \left(\nabla f(x(t)) + \beta A^T(Ax(t) - b) + A^T\left(\lambda(t) + \frac{t}{\alpha}\dot{\lambda}(t)\right)\right) \\
\qquad + \frac{\alpha\zeta}{t}\nabla^2\psi^*\left(\nabla\psi\left(x(t) + \frac{t}{\alpha}\dot{x}(t)\right)\right)\dot{x} = 0, \\
\ddot{\lambda}(t) + \frac{\alpha+1}{t}\dot{\lambda}(t) - A\left(x(t) + \frac{t}{\alpha}\dot{x}(t)\right) + b + \frac{\alpha\zeta}{t}\dot{\lambda}(t) = 0, \\
x(t_0) = x_0 \in \mathcal{X}, \dot{x}(t_0) = \dot{x}_0, \lambda(t_0) = \lambda_0, \dot{\lambda}(t_0) = \dot{\lambda}_0,
\end{cases}
\tag{15}
$$

i.e.,

$$
\begin{cases}
\ddot{x}(t) + \left(\frac{\alpha+1}{t}I_n + \frac{\alpha\zeta}{t}\nabla^2\psi^*\left(\nabla\psi\left(x(t) + \frac{t}{\alpha}\dot{x}(t)\right)\right)\right)\dot{x}(t) \\
\qquad + \nabla^2\psi^*\left(\nabla\psi\left(x(t) + \frac{t}{\alpha}\dot{x}(t)\right)\right)\nabla_x\mathcal{L}_\beta\left(x(t), \lambda(t) + \frac{t}{\alpha}\dot{\lambda}(t)\right) = 0, \\
\ddot{\lambda}(t) + \left(\frac{\alpha+1}{t}I_n + \frac{\alpha\zeta}{t}I_n\right)\dot{\lambda}(t) - \nabla_\lambda\mathcal{L}_\beta\left(x(t) + \frac{t}{\alpha}\dot{x}(t), \lambda(t)\right) = 0, \\
x(t_0) = x_0 \in \mathcal{X}, \dot{x}(t_0) = \dot{x}_0, \lambda(t_0) = \lambda_0, \dot{\lambda}(t_0) = \dot{\lambda}_0,
\end{cases}
\tag{16}
$$

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 $x(t)$ is inside the feasible set $\mathcal{X}$ in the intuitive understanding (the rigorous proof is postponed to **Lemma 8**). The form in (16) allows us to understand more intuitively why the proposed (14) is called the primal-dual method that is with primal variable x and dual variable λ.

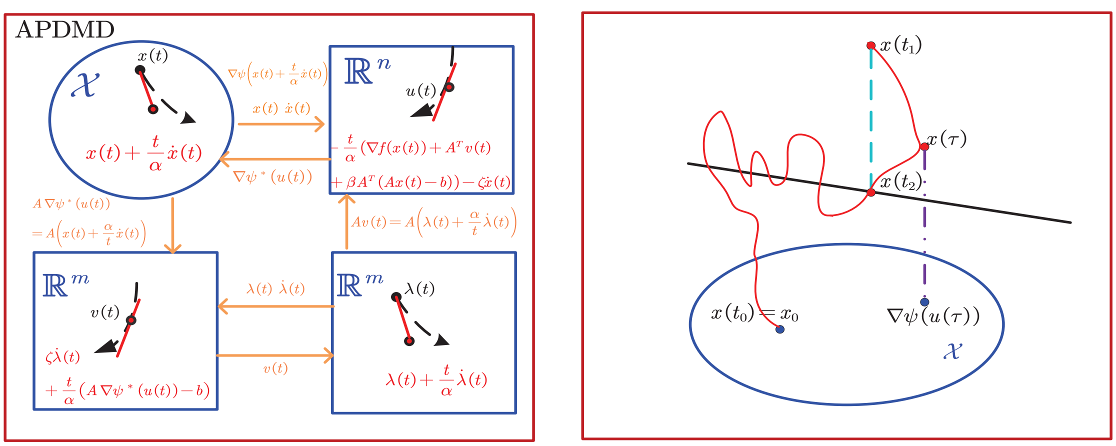

Figure 2: (left) Illustration of APDMD (14). (right) Feasibility of $x(t)$ to APDMD (14).

3.2 Interpretation of the APDMD

The APDMD (14) (i.e., (15)) can be interpreted from different perspectives: neurodynamic approach, Hamilton's system, game theory and control theory.

- **Neurodynamic approach perspective**. The APDMD (14) be regarded as a neurodynamic approach (describe the dynamic behavior of neurons), thus APDMD can

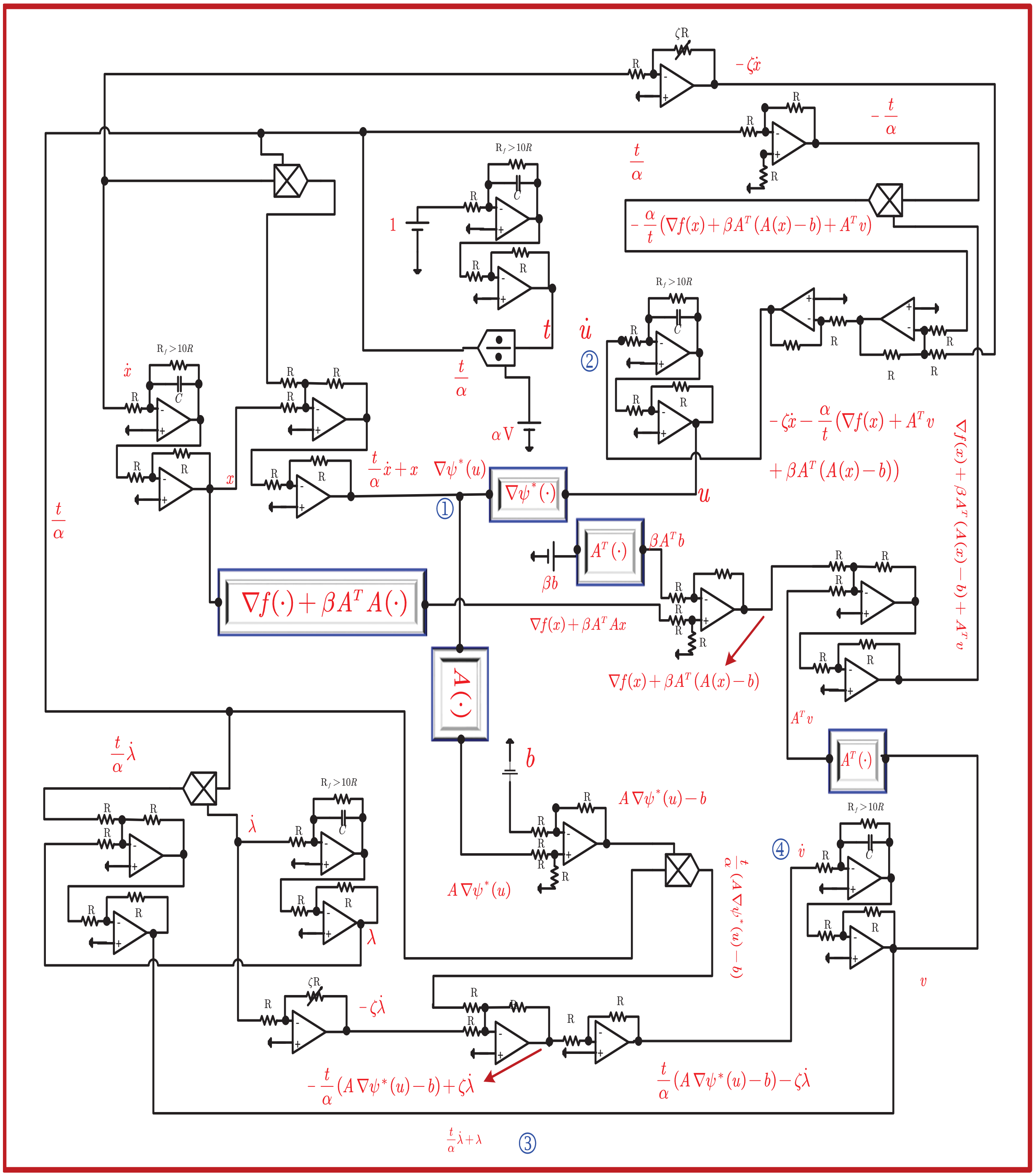

Figure 3: Circuit architecture diagram of APDMD (14) with ① $\frac{t}{\alpha}\dot{x}+x=\nabla\psi^*(u)$, ② $\dot{u}=-\frac{t}{\alpha}\left(\nabla f(x)+\beta A^T(Ax-b)+A^T v\right)-\zeta\dot{x}$, ③ $\frac{t}{\alpha}\dot{\lambda}+\lambda=v$, ④ $\dot{v}=\frac{t}{\alpha}(A\nabla\psi^*(u)-b)-\zeta\dot{\lambda}$.

be implemented by analog circuit like the Hopfield neural network in Hopfield and Tank (1986), as shown in Figure 3. The circuit in Figure 3 is designed according to APDMD (14) 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 3 is the stable point to the APDMD (14) (i.e., an optimal solution to the problem (1)). For more information about neurodynamic approaches and their circuits, please see (Kennedy and Chua (1988)).

- **Hamiltonian system perspective**. The Hamiltonian-based system approach is used to design accelerated dynamical approaches for solving the problem (3), which has recently been widely studied in (Diakonikolas and Jordan (2021); Wibisono et al. (2016)), and it plays a key role in designing of APDMD (14) and the development of our Lyapunov analysis method.

  Designing the first Hamiltonian (time-dependent) system

$$
\begin{aligned}
& H_1\left(\bar{x}(t), u(t), \eta(t), \dot{\bar{x}}(t), \dot{\eta}(t)\right) \\
=& \psi^*\left(u(t)\right) + \mathit{\Gamma}\left(\eta(t)\right)\left(f\left(\frac{\bar{x}(t)}{\eta(t)}\right) + \beta\eta(t)\left\|A\left(\frac{\bar{x}(t)}{\eta(t)}\right) - b\right\|^2\right. \\
& \left. + v(t)^T\left(A\left(\frac{\bar{x}(t)}{\eta(t)}\right) - b\right)\right) + \frac{\zeta\dot{\eta}(t)}{\eta^3(t)}\left(\dot{\bar{x}}(t)^T\bar{x}(t) - \frac{\dot{\eta}(t)\eta(t)}{2}\left\|\frac{\bar{x}(t)}{\eta(t)}\right\|^2\right)
\end{aligned}
$$

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

  The corresponding continuous-time dynamics of H_1 is

$$
\begin{aligned}
\frac{d}{dt}\left(\bar{x}(t)\right) =& \dot{\eta}(t)\frac{d\bar{x}(t)}{d\eta(t)} = \dot{\eta}(t)\nabla_u H_1\left(\bar{x}(t), u(t), \eta(t), \dot{\bar{x}}(t), \dot{\eta}(t)\right) = \dot{\eta}(t)\nabla\psi^*\left(u(t)\right) \\
& \Rightarrow \dot{x}(t) = \frac{\alpha}{t}\left(\nabla\psi^*\left(u(t)\right) - x(t)\right);
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}u(t) =& \dot{\eta}(t)\frac{du(t)}{d\eta(t)} = -\mathit{\Gamma}\left(\eta(t)\right)\dot{\eta}(t)\nabla_{\bar{x}}H_1\left(\bar{x}(t), u(t), \eta(t), \dot{\bar{x}}(t), \dot{\eta}(t)\right) \\
& \Rightarrow \dot{u}(t) = -\frac{t}{\alpha}\left(\nabla f\left(x(t)\right) + \beta A^T\left(Ax\left(x(t)\right) - b\right) + A^T v(t)\right) - \zeta\dot{x}(t).
\end{aligned}
$$

  Furthermore, setting the second Hamiltonian (time-dependent) system be

$$
\begin{aligned}
H_2\left(\bar{\lambda}(t), v(t), \eta(t), \dot{\bar{\lambda}}(t), \dot{\eta}(t)\right) =& \frac{1}{2}\|v(t)\|^2 - \left(\frac{\bar{\lambda}(t)}{\eta(t)}\right)^T\left(A\psi^*\left(u(t)\right) - b\right) \\
& + \frac{\zeta\dot{\eta}(t)}{\eta^3(t)}\left(\bar{\lambda}(t)^T\dot{\bar{\lambda}}(t) - \frac{1}{2}\dot{\eta}(t)\eta(t)\left\|\frac{\bar{\lambda}(t)}{\eta(t)}\right\|^2\right)
\end{aligned}
$$

  where $\bar{\lambda}(t) = \eta(t)\lambda(t)$.

The corresponding continuous-time dynamics of H_2 is

$$\begin{aligned}\frac{d}{dt}\left(\bar{\lambda}(t)\right) =& \dot{\eta}(t)\frac{d\bar{\lambda}(t)}{d\eta(t)} = \dot{\eta}(t)\nabla_v H_2\left(\bar{\lambda}(t), v(t), \eta(t), \dot{\bar{\lambda}}(t), \dot{\eta}(t)\right) = \dot{\eta}(t)v(t)\\ &\Rightarrow \dot{\lambda}(t) = \frac{\alpha}{t}(v(t)-\lambda(t))\\ \frac{d}{dt}v(t) =& \dot{\eta}(t)\frac{dv(t)}{d\eta(t)} = -\varGamma(\eta(t))\dot{\eta}(t)\nabla_{\bar{\lambda}}H_2\left(\bar{\lambda}(t), v(t), \eta(t), \dot{\bar{\lambda}}(t), \dot{\eta}(t)\right)\\ &\Rightarrow \dot{v}(t) = -\frac{t}{\alpha}(b-Au(t)) - \zeta\dot{\lambda}(t).\end{aligned}$$

From the above conclusions, the APDMD (14) can be obtained directly.

- **Game theoretic standpoint**. 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 (15). In the coupling term, the player $\lambda(t)$ takes account of the anticipated position of the player $x(t)$, i.e., $x(t)+\frac{t}{\alpha}\dot{x}(t)$. Conversely, for player $x(t)$, it takes account of the anticipated position of the player $\lambda(t)$, i.e., $\lambda(t)+\frac{t}{\alpha}\dot{\lambda}(t)$.

- **Control theoretic view**. The APDMD (15) can also be associated with control theory and state derivative feedback. Let $\chi(t)=\mathrm{col}(x(t),\lambda(t))$, the APDMD (15) can be written in the following formulation

$$\ddot{\chi}(t)+\frac{\alpha+1}{t}\dot{\chi}(t)=\varUpsilon(t,\chi(t),\dot{\chi}(t)),$$

with an operator $\varUpsilon$, 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 in (Michiels et al. (2009)).

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 (14) by the Cauchy-Lipschitz-Picard theorem in (Bolte (2003)).

Lemma 8. *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 (Krichene et al. (2016)), 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 ω, $a\in\mathbb{R}^n$ such that $(x(t_1)-a)^T\omega>0$ and $(x(t)-a)^T\omega<0, \forall x\in\mathcal{X}$. Let $d(x(t))=(x(t)-a)^T\omega$. Note that the trajectory $x(t)$ is continuous, then $t\to d(x(t))$ is continuous, and $d(x(t_0))=(x(t_0)-a)^T\omega<0$ due to $x_0\in\mathcal{X}$, $d(x(t_1))=(x(t_1)-a)^T\omega>0$ due to $x(t_1)\notin\mathcal{X}$. Thus, there is a t_2 such that $d(x(t_2))=0$, i.e., $d(x(t))>0$, $t\in[t_2,t_1)$, which implies t_2 is the last time when $x(t)$

crosses the separating hyperplane, and $t_2 = \inf\left\{t : d\left(x\left(t'\right)\right) \geqslant 0, \forall t' \in [t, t_1]\right\}$. We have $d(x(t_1)) - d(x(t_2)) > 0$, according to Taylor's theorem and $\dot{x}(t) = \frac{\alpha}{t}(\nabla\psi^*(u(t)) - x(t))$, there exists $\tau \in [t_2, t_1]$, such that

$$\begin{aligned} d(x(t_1)) - d(x(t_2)) =& \dot{d}(x(\tau)) \ = (\dot{x}(\tau))^T \omega = \frac{\alpha}{\tau}\omega^T (\nabla\psi^*(u(\tau)) - x(\tau)) \\ =& \frac{\alpha}{\tau}(d(\nabla\psi^*(u(\tau))) - d(x(\tau))) < 0, \end{aligned}$$

since $\nabla\psi^*(u_0) \in \mathcal{X}$. It leads to a contradiction, which concludes the proof and is shown in Figure 2 (right). ■

Definition 9. *Let $x : [t_0, +\infty) \to \mathcal{X}$, $\lambda : [t_0, +\infty) \to \mathbb{R}^m$ with the corresponding product space structure $\mathcal{X} \times \mathbb{R}^m$ and $t_0 > 0$. (x, u, λ, v) is a strong global solution of APDMD* (14) *if it satisfies the following properties:*

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

(ii) $\begin{cases} \dot{x}(t) = \frac{\alpha}{t}(\nabla\psi^*(u(t)) - x(t)), \\ \dot{u}(t) = -\frac{t}{\alpha}\left(\nabla f(x(t)) + \beta A^T(Ax(t) - b) + A^T v(t)\right) - \zeta\dot{x}(t), \\ \dot{\lambda}(t) = \frac{\alpha}{t}(v(t) - \lambda(t)), \\ \dot{v}(t) = \frac{t}{\alpha}(A\nabla\psi^*(u(t)) - b) - \zeta\dot{\lambda}(t), \end{cases}$ *, $t \in [t_0, +\infty)$;*

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

A mapping $x : [t_0, +\infty) \to \mathbb{R}^n$ is called locally absoutely continuous if it is absolutely continuous on every compact interval $[t_0, \mathcal{T}]$ with $\mathcal{T} > t_0$. For the absolutely continuous function $x : [t_0, +\infty) \to \mathbb{R}^n$, it has the following equivalent characterizations:

(a) There exists an integrable function $\mathsf{x} : [t_0, T) \to R^n$, such that

$$x(t) = x(t_0) + \int_{t_0}^t \mathsf{x}(s)\, ds, \forall t \in [t_0, \mathcal{T}];$$

(b) x is a continuous function and its distribuional derivative is Lebesque integrable on the interval $[t_0, \mathcal{T}]$;

(c) For every $\varepsilon > 0$, there exists $\pi > 0$, such that for any finite family of intervals $I_k = (a_k, b_k)$ from $[t_0, \mathcal{T}]$, the following implication is valid:

$$\left[I_k \cap I_j = \oslash \text{ and} \sum_k |b_k - a_k| < \pi\right] \Rightarrow \left[\sum_k \|x(b_k) - x(a_k)\| < \varepsilon\right].$$

Theorem 10. *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:

$$\begin{cases} \dot{Y}(t) = F(t, Y(t)) \\ Y(t_0) = (x(t_0), u(t_0), \lambda(t_0), v(t_0)), \end{cases} \tag{17}$$

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

According to the Cauchy-Lipschitz-Picard theorem, there exists a unique strong global solution for APDMD (14), i.e., (15), if the following conditions **(I)** and **(II)** are fulfilled.

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

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

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

$$\begin{aligned}&\left\|F(t,Y(t))-F\left(t,\hat{Y}(t)\right)\right\|\\ \le&\left(\frac{t^2}{\alpha^2}\left(\left(4\beta^2\delta_{\max}\left(A^TA\right)^2+4\mathfrak{l}_f^2+\delta_{\max}\left(A^TA\right)\mathfrak{l}_{\psi^*}^2+2\delta_{\max}\left(A^TA\right)\right)+\frac{\alpha^2\left(6+2\zeta+(2+\zeta)\mathfrak{l}_{\psi^*}^2\right)}{t^2}\right)^{\frac{1}{2}}\\ &\times\left\|Y(t)-\hat{Y}(t)\right\|,\end{aligned}$$

where $\delta_{\max}(A)$ is the maximum singular value of matrix A and the inequality holds from the Lipschitz continuous properties of ∇f and $\nabla\psi^*$ that have Lipschitz constants $\mathfrak{l}_f$ and $\mathfrak{l}_{\psi^*}$. Let $\mathfrak{l}(t)$ be $\left(\frac{t^2}{\alpha^2}\left(\left(4\beta^2\delta_{\max}\left(A^TA\right)^2+4\mathfrak{l}_f^2+\delta_{\max}\left(A^TA\right)\mathfrak{l}_{\psi^*}^2+2\delta_{\max}\left(A^TA\right)\right)+\frac{\alpha^2\left(6+2\zeta+(2+\zeta)\mathfrak{l}_{\psi^*}^2\right)}{t^2}\right)^{\frac{1}{2}}$, one has $\left\|F(t,Y(t))-F\left(t,\hat{Y}(t)\right)\right\|\le\mathfrak{l}(t)\left\|Y(t)-\hat{Y}(t)\right\|$.

Note that $\mathfrak{l}(t)$ is continuous on $[t_0,+\infty)$. Hence $\mathfrak{l}(\cdot)$ is integrable on $[t_0,\mathcal{T}]$ for all $0<t_0<\mathcal{T}<+\infty$.

The proof of **(II)**. For arbitrary $Y\in\mathcal{X}\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$ and $0<t_0<\mathcal{T}$, we have

$$\begin{aligned}&\int_{t_0}^{\mathcal{T}}\|F(t,Y(t))\|dt\\ =&\int_{t_0}^{\mathcal{T}}\Big(\tfrac{(1+\zeta)\alpha^2}{t^2}\|\nabla\psi^*(u(t))-x(t)\|^2+\tfrac{\alpha^2}{t^2}\|v(t)-\lambda(t)\|^2\\ &+\frac{t^2}{\alpha^2}\|A(\nabla\psi^*(u(t))-x^*)\|^2+\frac{t^2}{\alpha^2}\left\|\nabla f(x(t))+\beta A^TA(x(t)-x^*)-A^Tv(t)\right\|^2\Big)^{\frac{1}{2}}dt\\ \le&\int_{t_0}^{\mathcal{T}}\Big(\tfrac{2(1+\zeta)\alpha^2}{t^2}\|\nabla\psi^*(u(t))\|^2+\tfrac{4t^2\beta^2\delta_{\max}\left(A^TA\right)^2}{\alpha^2}\|x(t)\|^2\end{aligned}$$

$$
\begin{aligned}
&+\frac{2\alpha^2}{t^2}\|v(t)\|^2+\frac{2\alpha^2}{t^2}\|\lambda(t)\|^2+\frac{2(1+\zeta)\alpha^2}{t^2}\|x(t)\|^2+\frac{2t^2\delta_{\max}\left(A^TA\right)}{\alpha^2}\|\nabla\psi^*(u(t))\|^2\\
&+\frac{2t^2\delta_{\max}\left(A^TA\right)}{\alpha^2}\|x^*\|^2+\frac{4t^2}{\alpha^2}\|\nabla f(x(t))\|^2+\frac{4t^2\delta_{\max}\left(A^TA\right)}{\alpha^2}\|v(t)\|^2\\
&\quad+\frac{4t^2\beta^2\delta_{\max}\left(A^TA\right)^2}{\alpha^2}\|x^*\|^2\quad\Big)^{\frac{1}{2}}dt\\
\leq&\sqrt{\|\nabla\psi^*(u)\|^2+\|x\|^2+\|\lambda\|^2+\|v\|^2+\|x^*\|^2+\|\nabla f(x(t))\|^2}\\
&\times\left(\frac{(8+4\zeta)\alpha^2}{t^2}+\frac{2t^2}{\alpha^2}\left(4\delta_{\max}\left(A^TA\right)+4\beta^2\delta_{\max}\left(A^TA\right)^2+2\right)\right)^{\frac{1}{2}},
\end{aligned}
$$

and the conclusion holds by the continuity of the following function

$$
t\to\left(\frac{(8+4\zeta)\alpha^2}{t^2}+\frac{2t^2}{\alpha^2}\left(4\delta_{\max}\left(A^TA\right)+4\beta^2\delta_{\max}\left(A^TA\right)^2+2\right)\right)^{\frac{1}{2}}.
$$

In view of the above statements **(I)** and **(II)**, the existence and uniqueness of a strong global solution for the dynamical system (17) can be obtained. This leads directly to the existence and uniqueness of a strong global solution for APDMD. ■

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 (14) as follows.

Theorem 11. *Suppose **Assumption 3.1** holds. Let* $(x(t),\lambda(t))$ *and* (x^*,λ^*) *be the solution trajectory and optimal solution for APDMD* (14) *and problem* (1), *respectively. Then for any* $(x(t_0),u(t_0),\lambda(t_0),v(t_0))\in\mathcal{X}\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$, *and let* $\mathcal{L}_\beta(x(t),\lambda(t))=f(x(t))+\lambda(t)^T(Ax(t)-b)+\frac{\beta}{2}\|Ax(t)-b\|^2$, *and consider a Lyapunov function* $\mathcal{V}:[t_0,T)\to[0,+\infty)$ *which is given by*

$$
\mathcal{V}(t)=\mathcal{V}_1(t)+\mathcal{V}_2(t)+\mathcal{V}_3(t), \tag{18}
$$

where

$$
\begin{cases}
\mathcal{V}_1(t)=\frac{t^2}{\alpha^2}\left(\mathcal{L}_\beta(x(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda(t))\right)\\
\qquad\;\;=\frac{t^2}{\alpha^2}\left(f(x(t))-f(x^*)+\frac{\beta}{2}\|Ax(t)-b\|^2+A^T\lambda^*(x(t)-x^*)\right),\\
\mathcal{V}_2(t)=D_{\psi^*}(u(t),u^*)+\frac{\zeta}{2}\|x(t)-x^*\|^2\\
\qquad\;\;=\psi^*(u(t))-\psi^*(u^*)-\nabla\psi^*(u^*)^T(u(t)-u^*)+\frac{\zeta}{2}\|x(t)-x^*\|^2,\\
\mathcal{V}_3(t)=D_h(v(t),v^*)+\frac{\zeta}{2}\|\lambda(t)-\lambda^*\|^2\\
\qquad\;\;=\frac{1}{2}\|v(t)-v^*\|^2+\frac{\zeta}{2}\|\lambda(t)-\lambda^*\|^2,
\end{cases}
$$

where $D_{\psi^*}(u,u^*)$ *with* $u^*=\nabla\psi(x^*)$, $D_h(v,v^*)$ *with* $v^*=\lambda^*$ *are two Bregman divergences associated with suitable functions* ψ^* *and* h *that are determined by different requirements, respectively. The following statements are true:*

(I) $\zeta = 0, \beta \geq 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we have*

1) $\frac{t}{\alpha}\dot{x}(t) + x(t) = \psi^*(u(t))$ *and* $\frac{t}{\alpha}\dot{\lambda}(t) + \lambda(t) = v(t)$ *are bounded;*

2) One has

$$\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) \leq \frac{\alpha^2 \mathcal{V}(t_0)}{t^2}; \tag{19}$$

3) if, in addition, $\beta > 0$,

$$\|Ax(t) - b\|^2 \leq \frac{2\alpha^2 \mathcal{V}(t_0)}{\beta t^2}; \tag{20a}$$

$$\int_{t_0}^{+\infty} t\,\|Ax(t) - b\|^2 dt < +\infty; \tag{20b}$$

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we have*

$$\int_{t_0}^{+\infty} t\,(\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)))dt < +\infty; \tag{21}$$

(II) $\zeta > 0, \beta \geqslant 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *one has*

1) $x(t)$ *and* $\lambda(t)$ *are bounded;*

$$\|\dot{x}(t)\| = O\left(\frac{1}{t}\right), \left\|\dot{\lambda}(t)\right\| = O\left(\frac{1}{t}\right); \tag{22}$$

2)

$$\int_{t_0}^{+\infty} t\,\|\dot{x}(t)\|^2\, dt \leqslant \frac{\alpha}{\zeta}\mathcal{V}(t_0) < +\infty; \tag{23a}$$

$$\int_{t_0}^{+\infty} t\left\|\dot{\lambda}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\mathcal{V}(t_0) < +\infty; \tag{23b}$$

3)

$$\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) \leq \frac{\alpha^2 \mathcal{V}(t_0)}{t^2}; \tag{24a}$$

$$\|Ax(t) - b\| \leqslant \frac{2\zeta\alpha^2 C_1}{t^2}; \tag{24b}$$

$$|f(x(t)) - f(x^*)| \leqslant \frac{\alpha^2}{t^2}\left(\mathcal{V}(t_0) + 2\zeta\|\lambda^*\| C_1\right) + \frac{2\beta\zeta^2\alpha^4 C_1^2}{t^4}; \tag{24c}$$

where $C_1 = \sup\limits_{t\in[t_0,+\infty)} \left\{ \frac{t}{\zeta\alpha}\left\|\dot{\lambda}(t)\right\| + \frac{t_0}{\zeta\alpha}\left\|\dot{\lambda}(t_0)\right\| + \frac{1+\zeta}{\zeta}\|\lambda(t) - \lambda(t_0)\| + \frac{t_0^2}{\alpha^2}\|Ax(t_0) - b\| \right\} < +\infty$.

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we have*

$$\int_{t_0}^{+\infty} t\,(\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)))dt < +\infty. \tag{25}$$

Proof Note that the Laypunov function $\mathcal{V}_1(t) \geqslant \frac{\beta t^2}{2\alpha^2}\|Ax(t)-b\|^2 \geqslant 0$ since

$$\begin{aligned}
&\mathcal{L}_\beta(x(t),\lambda^*) - \mathcal{L}_\beta(x^*,\lambda(t)) \\
=&f(x(t)) - f(x^*) + \frac{\beta}{2}\|Ax(t)-b\|^2 + A^T\lambda^*(x(t)-x^*) \\
\in&f(x(t)) - f(x^*) + \frac{\beta}{2}\|Ax(t)-b\|^2 - (\nabla f(x^*) + \mathcal{N}_\mathcal{X}(x^*))^T(x(t)-x^*) \\
\geq&f(x(t)) - f(x^*) - f(x(t)) + f(x^*) + \frac{\beta}{2}\|Ax(t)-b\|^2 \\
\geq&\frac{\beta}{2}\|Ax(t)-b\|^2 \geq 0, \forall (x(t),\lambda(t)) \in \mathcal{X}\times\mathbb{R}^m,
\end{aligned}$$

where $\mathcal{N}_\mathcal{X}(x^*) = \left\{w \in \mathbb{R}^n | w^T(v-x^*) \leq 0, \forall v \in \mathcal{X}\right\}$ is normal cone to set Ω at point x^*.

When $\zeta = 0$. Since $\psi^*(\cdot)$ is convex, it gives $D_{\psi^*}(u(t),u^*) \geq 0$, and $h(\cdot) = \frac{1}{2}\|\cdot\|^2$ is strongly convex, we can obtain that $D_h(v(t),\lambda^*) \geq 0$ and $\|v(t)\| \to +\infty$, $\mathcal{V}_2(t) \to +\infty$, i.e., $\mathrm{V}_2(t)$ is radially unbounded of variable $v(t)$. Thus, $\mathcal{V}(t)$ is nonnegative and radially unbounded of variable $v(t)$.

When $\zeta > 0$. According to $\frac{\zeta}{2}\|x(t)-x^*\|^2$, one has $\|x(t)\| \to +\infty, \mathcal{V}_2(t) \to +\infty$ which means $\mathcal{V}_2(t)$ is nonnegative and radially unbounded of $x(t)$. In addition, for the $\mathcal{V}_3(t)$, since $D_h(v(t),v^*) = \frac{1}{2}\|v(t)-v^*\|^2$ and $\frac{\zeta}{2}\|\lambda(t)-\lambda^*\|^2 \geqslant 0$ are strongly convex functions of variables $v(t)$ and $\lambda(t)$, it implies $\|v(t)\| \to +\infty$ or $\|\lambda(t)\| \to +\infty$, $\mathcal{V}_3(t) \to +\infty$. Thus, $\mathcal{V}_3(t)$ is nonnegative and radially unbounded of variables $v(t)$ and $\lambda(t)$. Thus, $\mathcal{V}(t)$ is nonnegative and radially unbounded of variables $x(t)$, $v(t)$ and $\lambda(t)$.

The derivatives of $\mathcal{V}_1(t)$, $\mathcal{V}_2(t)$ and $\mathcal{V}_3(t)$ along the trajectory of APDMD with $x(t) \in \mathcal{X}$, $\nabla\psi^*(u^*) = x^*$ and $b = Ax^*$ are given by

$$\begin{aligned}
\dot{\mathcal{V}}_1(t) =&\frac{2t}{\alpha^2}\left(\mathcal{L}_\beta(x(t),\lambda^*) - \mathcal{L}_\beta(x^*,\lambda(t))\right) + \left(\nabla\psi^*(u(t)) - x(t)\right)^T \\
&\times\frac{t}{\alpha}\left(\nabla f(x(t)) + \frac{\beta}{2}A^T(Ax(t)-b) + A^T\lambda^*\right) \\
=&\frac{2t}{\alpha^2}\left(f(x(t)) - f(x^*) + \frac{\beta}{2}\|Ax(t)-b\|^2 + \left(A^T\lambda^*\right)^T(x(t)-x^*)\right) \\
&+\frac{t}{\alpha}\left(\nabla f(x(t)) + \beta A^TA(x(t)-x^*) + A^T\lambda^*\right)^T \\
&\times\left(\nabla\psi^*(u(t)) - \nabla\psi^*(u^*)\right) + \left(\nabla\psi^*(u^*) - x(t)\right)^T \\
&\times\frac{t}{\alpha}\left(\nabla f(x(t)) + \beta A^TA(x(t)-x^*) + A^T\lambda^*\right),
\end{aligned} \tag{26}$$

$$\begin{aligned}
\dot{\mathcal{V}}_2(t) =&\zeta(x(t)-x^*)^T\dot{x}(t) - \frac{t}{\alpha}\left(\nabla\psi^*(u(t)) - \nabla\psi^*(u^*)\right)^T \\
&\times\left(\nabla f(x(t)) + \beta A^T(Ax(t)-b) + A^Tv(t)\right) \\
&-\zeta\left(\nabla\psi^*(u(t)) - \nabla\psi^*(u^*)\right)^T\dot{x}(t) \\
=&-\frac{\zeta t}{\alpha}\|\dot{x}(t)\|^2 - \frac{t}{\alpha}\left(\nabla\psi^*(u(t)) - \nabla\psi^*(u^*)\right)^T \\
&\times\left(\nabla f(x(t)) + \beta A^TA(x(t)-x^*) + A^T\lambda^*\right) \\
&-\frac{t}{\alpha}\left(\nabla\psi^*(u(t)) - \nabla\psi^*(u^*)\right)^T\left(A^Tv(t) - A^T\lambda^*\right),
\end{aligned} \tag{27}$$

$$\begin{aligned}\dot{\mathcal{V}}_3(t) =&\frac{t}{\alpha}\left(v(t)-\lambda^*\right)^T A\left(\nabla\psi^*(u(t))-x^*\right)\\&-\zeta\left(v(t)-\lambda^*\right)^T\dot{\lambda}(t)+\zeta\left(\lambda(t)-\lambda^*\right)^T\dot{\lambda}(t)\\=&\frac{t}{\alpha}\left(v(t)-\lambda^*\right)^T A\left(\nabla\psi^*(u(t))-x^*\right)-\frac{\zeta t}{\alpha}\left\|\dot{\lambda}(t)\right\|^2.\end{aligned}\tag{28}$$

Adding (26), (27) and (28) together, and using $Ax^*=b$, we have

$$\begin{aligned}\dot{\mathcal{V}}(t) =&\dot{\mathcal{V}}_1(t)+\dot{\mathcal{V}}_2(t)+\dot{\mathcal{V}}_3(t)\\=&\frac{2t}{\alpha^2}\left(\mathcal{L}_\beta(x(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda(t))\right)-\frac{\zeta t}{\alpha}\|\dot{x}(t)\|^2-\frac{\zeta t}{\alpha}\left\|\dot{\lambda}(t)\right\|^2\\&+\frac{t}{\alpha}\left(x^*-x(t)\right)^T\left(\nabla f(x(t))+\beta A^TA(x(t)-x^*)+A^T\lambda^*\right)\\\leq&\frac{2t}{\alpha^2}\left(\mathcal{L}_\beta(x(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda(t))\right)-\frac{t}{\alpha}\left(\mathcal{L}_\beta(x^*,\lambda^*)-\mathcal{L}_\beta(x(t),\lambda^*)\right)\\&-\frac{\beta t}{2\alpha}\|Ax(t)-b\|^2-\frac{\zeta t}{\alpha}\|\dot{x}(t)\|^2-\frac{\zeta t}{\alpha}\left\|\dot{\lambda}(t)\right\|^2\\=&-t\frac{\alpha-2}{\alpha}\left(\mathcal{L}_\beta(x(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda(t))\right)-\frac{\beta t}{2\alpha}\|Ax(t)-b\|^2\\&-\frac{\zeta t}{\alpha}\|\dot{x}(t)\|^2-\frac{\zeta t}{\alpha}\left\|\dot{\lambda}(t)\right\|^2\leq 0,\end{aligned}\tag{29}$$

where the second equation holds from $\nabla\psi^*(u^*)=x^*$, the first inequality is satisfied due to the convexity of $\mathcal{L}_\beta(x(t),\lambda^*)$ of $x(t)$ with fixed λ^*, the third equation holds because of $\mathcal{L}_\beta(x^*,\lambda(t))=\mathcal{L}_\beta(x^*,\lambda^*)$ and the last inequality is established since $\alpha\geq 2$, $\zeta\geq 0$ and $\mathcal{L}_\beta(x(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda(t))\geq 0$.

From (29), we get $\dot{\mathcal{V}}(t)\leq 0, t\geq t_0>0$, which means $\mathcal{V}(t)$ is nonincreasing when $t\geq t_0>0$, i.e., $0<\mathcal{V}(t)\leqslant\mathcal{V}(t_0), t\geq t_0>0$.

(I) $\zeta=0,\beta\geq 0$.

(i) If $\alpha\geqslant 2$.

1) According to the definition of $\mathcal{V}(t)$ in (18) and $\dot{\mathcal{V}}(t)\leq 0, t\geq t_0>0$, one has

$$\begin{aligned}0\leq&\frac{1}{2\mathfrak{l}_{\psi^*}}\left\|\frac{t}{\alpha}\dot{x}(t)+x(t)-x^*\right\|^2+\frac{1}{2}\left\|\frac{t}{\alpha}\dot{\lambda}(t)+\lambda(t)-\lambda^*\right\|^2\\=&\frac{1}{2\mathfrak{l}_{\psi^*}}\|\psi^*(u(t))-x^*\|^2+\frac{1}{2}\|v(t)-\lambda^*\|^2\\\leq&D_{\psi^*}(u(t),u^*)+D_h(v(t),\lambda^*)\leqslant\mathcal{V}(t_0)<+\infty,\end{aligned}$$

which implies that $\sup_{t\in[t_9,+\infty)}\left\|\frac{t}{\alpha}\dot{\lambda}(t)+\lambda(t)\right\|<+\infty$ and $\sup_{t\in[t_9,+\infty)}\left\|\frac{t}{\alpha}\dot{x}(t)+x(t)\right\|<+\infty, \forall x(t)\in\mathcal{X}$, i.e., $\frac{t}{\alpha}\dot{x}(t)+x(t)=\psi^*(u(t))$ and $\frac{t}{\alpha}\dot{\lambda}(t)+\lambda(t)=v(t)$ are bounded.

2) From the definition of $\mathcal{V}(t)$ in (18) and $\dot{\mathcal{V}}(t)\leq 0, t\geq t_0>0$, one also has

$$0\leqslant\frac{t^2}{\alpha^2}\left(\mathcal{L}_\beta(x(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda(t))\right)\leqslant\mathcal{V}(t_0)<+\infty,$$

which means $\mathcal{L}_\beta(x(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda(t))\leq\frac{\alpha^2\mathcal{V}(t_0)}{t^2},\ t\geqslant t_0>0, x(t)\in\mathcal{X}$.

3) From the definition of $\mathcal{V}(t)$ in (18) and $\dot{\mathcal{V}}(t) \le 0, t \ge t_0 > 0$, we obtain

$$0 \leqslant \frac{\beta t^2}{2\alpha^2} \|Ax(t) - b\|^2 \leqslant \mathcal{V}(t_0) < +\infty,$$

which implies $\|Ax(t) - b\|^2 \le \frac{2\alpha^2 \mathcal{V}(t_0)}{\beta t^2}, \forall\, t \geqslant t_0, x(t) \in \mathcal{X}$. In addition, (29) with $\alpha \geqslant 2, \zeta = 0, \beta > 0$, we can get $\dot{\mathcal{V}}(t) \leqslant -\frac{\beta t}{2\alpha} \|Ax(t) - b\|^2, \forall\, t \geqslant t_0, x(t) \in \mathcal{X}$, and integrating the inequality above yields

$$\int_{t_0}^{+\infty} t \|Ax(t) - b\|^2 dt \leqslant \frac{2\alpha}{\beta}\mathcal{V}(t_0) - \frac{2\alpha}{\beta}\mathcal{V}(+\infty) \leqslant \frac{2\alpha}{\beta}\mathcal{V}(t_0) < +\infty, \forall x(t) \in \mathcal{X}.$$

(ii) $\alpha > 2$. From (29) with $\alpha > 2, \zeta = 0, \beta \ge 0$, we can get

$$\dot{\mathcal{V}}(t) \leqslant -t\frac{\alpha - 2}{\alpha} \left(\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t))\right), \forall\, t \geqslant t_0, x(t) \in \mathcal{X},$$

and integrating the inequality above gives

$$\begin{aligned} &\int_{t_0}^{+\infty} t\left(\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t))\right)dt \\ \leqslant &\frac{\alpha}{\alpha - 2}\mathcal{V}(t_0) - \frac{\alpha}{\alpha - 2}\mathcal{V}(+\infty) \leqslant \frac{\alpha}{\alpha - 2}\mathcal{V}(t_0) < +\infty, \forall x(t) \in \mathcal{X}. \end{aligned}$$

(II) $\zeta > 0, \beta \geqslant 0$.

(i) $\alpha \geqslant 2$.

1) From the definition of $\mathcal{V}(t)$ in (18) and $\dot{\mathcal{V}}(t) \le 0, t \ge t_0 > 0$, one has

$$\begin{aligned} &\sup_{t\in[t_0,+\infty)} \left\|\frac{t}{\alpha}\dot{x}(t) + x(t) - x^*\right\| \leqslant \sqrt{2\mathcal{V}(t_0)} < +\infty, \forall x(t) \in \mathcal{X} \\ &\sup_{t\in[t_0,+\infty)} \left\|\frac{t}{\alpha}\dot{\lambda}(t) + \lambda(t) - \lambda^*\right\| \leqslant \sqrt{2\mathcal{V}(t_0)} < +\infty, \\ &\sup_{t\in[t_0,+\infty)} \|x(t) - x^*\| \leqslant \sqrt{\frac{2}{\zeta}\mathcal{V}(t_0)} < +\infty, \forall x(t) \in \mathcal{X} \\ &\sup_{t\in[t_0,+\infty)} \|\lambda(t) - \lambda^*\| \leqslant \sqrt{\frac{2}{\zeta}\mathcal{V}(t_0)} < +\infty, \end{aligned}$$

which means that $x(t), \forall\, t \geqslant t_0, x(t) \in \mathcal{X}$ and $\lambda(t)$ are bounded. In addition, by using the triangle inequality, we obtain

$$\begin{aligned} t\|\dot{x}(t)\| \leqslant &\alpha \left\|\frac{t}{\alpha}\dot{x}(t) + x(t) - x^*\right\| + \alpha\|x(t) - x^*\| \\ \leqslant &a\left(\sqrt{\zeta} + 1\right)\sqrt{\frac{2}{\zeta}\mathcal{V}(t_0)} < +\infty, \forall\, t \geqslant t_0, x(t) \in \mathcal{X}; \\ t\left\|\dot{\lambda}(t)\right\| \leqslant &\alpha \left\|\frac{t}{\alpha}\dot{\lambda}(t) + \lambda(t) - \lambda^*\right\| + \alpha\|\lambda(t) - \lambda^*\| \\ \leqslant &a\left(\sqrt{\zeta} + 1\right)\sqrt{\frac{2}{\zeta}\mathcal{V}(t_0)} < +\infty, \forall\, t \geqslant t_0, \end{aligned}$$

it means $\|\dot{x}(t)\| = O\left(\frac{1}{t}\right), \forall t \geqslant t_0, x(t) \in \mathcal{X}$ and $\left\|\dot{\lambda}(t)\right\| = O\left(\frac{1}{t}\right), \forall t \geqslant t_0$.

2) From (29) with $\alpha \geq 2, \zeta > 0, \beta \geq 0$, we can get $\dot{\mathcal{V}}(t) \leqslant -\frac{\zeta t}{\alpha}\|\dot{x}(t)\|^2 - \frac{\zeta t}{\alpha}\left\|\dot{\lambda}(t)\right\|^2, \forall\, t \geqslant t_0, x(t) \in \mathcal{X}$, and integrating the inequality above, we can obtain for any $t \geq t_0 > 0$

$$\int_{t_0}^{+\infty} t\,\|\dot{x}(t)\|^2\,dt \leqslant \frac{\alpha}{\zeta}\mathcal{V}(t_0) < +\infty,$$
$$\int_{t_0}^{+\infty} t\left\|\dot{\lambda}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\mathcal{V}(t_0) < +\infty.$$

3) In addition, together with (18) and $\dot{\mathcal{V}}(t) \leq 0, t \geq t_0 > 0$, we have

$$0 \leqslant \frac{t^2}{\alpha^2}\left(\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t))\right) \leqslant \mathcal{V}(t_0) < +\infty, \tag{30}$$

which implies $\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) \leq \frac{\alpha^2\mathcal{V}(t_0)}{t^2}$.

From the first, third and fourth equalities in APDMD (14), for every $t \geq t_0$, we have

$$\begin{aligned}
\lambda(t) - \lambda(t_0) =& \int_{t_0}^{t} \dot{\lambda}(s)\,ds \\
=& \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{s}{\alpha}\left(A\nabla\psi^*(u(s)) - b\right) ds \;\; - \frac{1}{\zeta}\int_{t_0}^{t} \dot{v}(s)\,ds \\
=& \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{s}{\alpha}\left(A\nabla\psi^*(u(s)) - b\right) ds \;\; + \frac{t_0}{\zeta\alpha}\dot{\lambda}(t_0) + \frac{1}{\zeta}\lambda(t_0) - \frac{t}{\zeta\alpha}\dot{\lambda}(t) - \frac{1}{\zeta}\lambda(t) \\
\Rightarrow& \frac{t}{\zeta\alpha}\dot{\lambda}(t) + \frac{1+\zeta}{\zeta}\lambda(t) - \frac{t_0}{\zeta\alpha}\dot{\lambda}(t_0) - \frac{1+\zeta}{\zeta}\lambda(t_0) \\
=& \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{s}{\alpha}\left(Ax(s) - b\right) ds \;\; + \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{s^2}{\alpha^2} d\left(Ax(s) - b\right) ds \\
=& \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{s}{\alpha}\left(Ax(s) - b\right) ds \;\; + \frac{t^2}{\zeta\alpha^2}\left(Ax(t) - b\right) \\
& - \frac{t_0^2}{\zeta\alpha^2}\left(Ax(t_0) - b\right) - \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{2s}{\alpha^2}\left(Ax(s) - b\right) ds \\
=& \frac{t^2}{\zeta\alpha^2}\left(Ax(t) - b\right) - \frac{t_0^2}{\zeta\alpha^2}\left(Ax(t_0) - b\right) + \frac{1}{\zeta}\int_{t_0}^{t} s\left(\frac{\alpha-2}{\alpha^2}\right)\; \left(Ax(s) - b\right) ds,
\end{aligned} \tag{31}$$

where the second equality holds from the fourth equality in APDMD (14) $\dot{\lambda}(t) = \frac{t}{\zeta\alpha}$ $\left(A\nabla\psi^*(u(t)) - b\right) - \frac{1}{\zeta}\dot{v}(t)$, the third equality is satisfied due to $v(t) = \frac{t}{\alpha}\dot{\lambda}(t) + \lambda(t)$, i.e., the third equality in APDMD (14), and the fourth equality holds because of $\nabla\psi^*(u(t)) = x(t) + \frac{t}{\alpha}\dot{x}(t)$, i.e., the first equality in APDMD (14).

It follows from (31) that,

$$\left\|\frac{t^2}{\zeta\alpha^2}\left(Ax(t) - b\right) + \int_{t_0}^{t}\left(\frac{\alpha-2}{s}\right)\; \tfrac{s^2}{\zeta\alpha^2}\left(Ax(s) - b\right) ds\;\right\| \leqslant C_1, \forall t \geqslant t_0 > 0, x(t) \in \mathcal{X}. \tag{32}$$

where $C_1 = \sup\limits_{t\in[t_0,+\infty)} \left\{ \frac{t}{\zeta\alpha} \left\| \dot{\lambda}(t) \right\| + \frac{t_0}{\zeta\alpha} \left\| \dot{\lambda}(t_0) \right\| + \frac{1+\zeta}{\zeta} \left\| \lambda(t) - \lambda(t_0) \right\| + \frac{t_0^2}{\alpha^2} \left\| Ax(t_0) - b \right\| \right\} < +\infty$. Setting

$$\mathsf{g}(t) = \frac{t^2}{\zeta\alpha^2} \left(Ax(t) - b \right), \mathsf{a}(t) = \frac{\alpha-2}{\alpha^2}, \forall t \geqslant t_0 > 0,$$

and applying **Lemma 12** to get that

$$\begin{aligned} &\frac{t^2}{\zeta\alpha^2} \left\| Ax(t) - b \right\| \leqslant 2C_1 \\ &\Rightarrow \left\| Ax(t) - b \right\| \leqslant \frac{2\zeta\alpha^2 C_1}{t^2}, \forall t \geqslant t_0 > 0, x(t) \in \mathcal{X}. \end{aligned} \tag{33}$$

Since $\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) = f(x(t)) - f(x^*) + \frac{\beta}{2} \left\| Ax(t) - b \right\|^2 + (Ax(t) - b)^T \lambda^*$. Using (30) and (33) we get

$$\begin{aligned} &\left| f(x(t)) - f(x^*) \right| \\ \leqslant &\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) + \left\| \lambda^* \right\| \left\| Ax(t) - b \right\| + \frac{\beta}{2} \left\| Ax(t) - b \right\|^2 \\ \leqslant &\frac{\alpha^2}{t^2} \left(V(t_0) + 2\zeta \left\| \lambda^* \right\| C_1 \right) + \frac{2\beta\zeta^2\alpha^4 C_1^2}{t^4}, \forall t \geqslant t_0 > 0, x(t) \in \mathcal{X}. \end{aligned}$$

(ii) $\alpha > 2$. The inequality (29) with $\alpha > 2, \zeta > 0, \beta \geq 0$, we can get $\dot{\mathcal{V}}(t) \leqslant -t\frac{\alpha-2}{\alpha} \left(\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) \right), \forall\ t \geqslant t_0, x(t) \in \mathcal{X}$, and integrating the inequality above gives

$$\begin{aligned} \int_{t_0}^{+\infty} t \left(\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) \right) dt \leqslant &\frac{\alpha}{\alpha-2} \mathcal{V}(t_0) - \frac{\alpha}{\alpha-2} \mathcal{V}(+\infty) \\ \leqslant &\frac{\alpha}{\alpha-2} \mathcal{V}(t_0) < +\infty, \forall x(t) \in \mathcal{X}. \end{aligned}$$

Thus, the proof is completed. ■

Lemma 12. *(Hulett and Nguyen (2023), He et al. (2022), He et al. (2023)) Assume that* $\mathsf{g} : [t_0, +\infty) \to \mathbb{R}$ *is a continuously differentiable function of* t*, and* $\mathsf{a} : [t_0, +\infty) \to [0, +\infty)$ *is a continuous function,* $t \geqslant t_0 > 0, C \geqslant 0$*. If*

$$\left| \mathsf{g}(t) + \int_{t_0}^{t} \mathsf{a}(s)\, \mathsf{g}(s)\, ds \right| \leqslant C, t \geqslant t_0 > 0,$$

then

$$\sup_{t \geqslant t_0} \left| \mathsf{g}(t) \right| \leqslant 2C < +\infty, t \geqslant t_0 > 0.$$

3.5 Examples of the APDMD

In this subsection, we will illustrate some examples of the APDMD (14), i.e., (15), in other words, the APDMD can reduce to some classical and new dynamical approaches when choosing different constraint set $\mathcal{X}$, i.e., $\mathcal{X}$ is an Euclidean space, $\mathcal{X}$ is a positive-orthant constrained set, $\mathcal{X}$ is a unit simplex set and $\mathcal{X}$ is a closed and convex set, and its projection operator $P_{\mathcal{X}}$ has a closed form solution (the detailed description is given in (Parikh et al. (2014)).

Case 1: If $\mathcal{X}$ is an Euclidean space, i.e., $\mathcal{X}=\mathbb{R}^n$. Let $\psi\left(x(t)+\frac{\alpha}{t}\dot{x}(t)\right)=\frac{1}{2}\left\|x(t)+\frac{\alpha}{t}\dot{x}(t)\right\|^2$, then, one has $\nabla\psi\left(x(t)+\frac{\alpha}{t}\dot{x}(t)\right)=u(t)$, $\psi^*(u(t))=\frac{1}{2}\|u(t)\|^2$, $\nabla\psi^*(u(t))=u(t)$, $\nabla^2\psi^*\left(\nabla\psi\left(x(t)(t)+\frac{\alpha}{t}\dot{x}(t)\right)\right)\equiv I_n$. The APDMD (15) reduces to the classical accelerated primal-dual dynamical approaches in (He et al. (2021); Zeng et al. (2022); Boţ and Nguyen (2021); Attouch et al. (2022)).

$$\begin{cases}\ddot{x}(t)+\frac{\alpha+1+\alpha\zeta}{t}\dot{x}(t)\\ \qquad +\left(\nabla f(x(t))+\beta A^T(Ax(t)-b)+A^T\left(\lambda(t)+\frac{t}{\alpha}\dot{\lambda}(t)\right)\right)=0,\\ \ddot{\lambda}(t)+\frac{\alpha+1+\alpha\zeta}{t}\dot{\lambda}(t)-A\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)+b=0,\\ x(t_0)=x_0\in\mathbb{R}^n,\dot{x}(t_0)=0,\lambda(t_0)=\lambda_0,\dot{\lambda}(t_0)=0.\end{cases}\tag{34}$$

Case 2: In (Banerjee et al. (2005)), if $\mathcal{X}$ is a positive-orthant constrained set, i.e., $\mathcal{X}=\mathbb{R}^n_+$. 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 $\nabla\psi(x)=\mathrm{col}(1+\ln x_1,...,1+\ln x_n)$ and $\psi^*(u)=\sum_{i=1}^n e^{u_i-1}$, $\nabla\psi^*(u)=\mathrm{col}\left(e^{u_1-1},...,e^{u_n-1}\right)$, $\nabla^2\psi^*(u(t))=\mathrm{diag}\left(e^{u_1-1},...,e^{u_n-1}\right)$. Thus, for all $x\in\mathbb{R}^n_+$, we have

$$\begin{aligned}&\nabla^2\psi^*\left(\nabla\psi\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)\right)\\ =&\mathrm{diag}\left(x_1(t)+\frac{t}{\alpha}x_1(t),...,x_n(t)+\frac{t}{\alpha}x_n(t)\right)=\mathrm{diag}\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right).\end{aligned}$$

Correspondingly, APDMD (15) is turned into

$$\begin{cases}\ddot{x}(t)+\frac{\alpha+1}{t}\left(I_n+\frac{\alpha\zeta}{\alpha+1}\mathrm{diag}\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)\right)\dot{x}(t)\\ +\mathrm{diag}\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)\left(\nabla f(x(t))+\beta A^T(Ax(t)-b)+A^T\left(\lambda(t)+\frac{t}{\alpha}\dot{\lambda}(t)\right)\right)=0,\\ \ddot{\lambda}(t)+\frac{\alpha+1+\alpha\zeta}{t}\dot{\lambda}(t)-A\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)+b=0,\\ x(t_0)=x_0\in\mathcal{X},\dot{x}(t_0)=\dot{x}_0,\lambda(t_0)=\lambda_0,\dot{\lambda}(t_0)=\dot{\lambda}_0,\end{cases}\tag{35}$$

In addition, for the constrained set $\mathcal{X}=\mathbb{R}^n_+$, the distance function can also be selected as $\psi(x)=-\sum_{i=1}^n\ln x_i$. Then, we have $\nabla\psi(x)=\mathrm{col}\left(-\frac{1}{x_1},...,-\frac{1}{x_n}\right)$ and the Bregman divergence of ψ 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 the **Itakura-Saito** divergence). Further we have $\psi^*(u)=-\sum_{i=1}^n(1+\ln(-u_i))$ on $\mathbb{R}^n_-$, $\nabla\psi^*(u)=$

$\mathrm{col}\left(-\frac{1}{u_1},...,-\frac{1}{u_n}\right)$ and $\nabla^2\psi^* = \mathrm{diag}\left(\frac{1}{(u_1)^2},...,\frac{1}{(u_n)^2}\right)$ on $\mathbb{R}^n_-$. To sum up, one has

$$\begin{aligned}&\nabla^2\psi^*\left(\nabla\psi\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)\right)\\ =&\mathrm{diag}\left(\left(x_1(t)+\frac{t}{\alpha}\dot{x}_1(t)\right)^2,...,\left(x_n+\frac{t}{\alpha}\dot{x}_n(t)\right)^2\right).\end{aligned}\tag{36}$$

Case 3: In (Beck and Teboulle (2003)), if $\mathcal{X}$ is a unit simplex set, i.e., $\mathcal{X}=\triangle=\left\{x\in\mathbb{R}^n_+ \middle| \sum_{i=1}^n x_i = 1\right\}$. Considering a distance-generating function $\psi(x)=\sum_{i=1}^n x_i\ln x_i + I_{\mathcal{X}}(x)$, where $\sum_{i=1}^n x_i\ln x_i$ is the negative entropy function and $I_{\mathcal{X}}(x)$ is the indicator function on $\mathcal{X}$ and its Bregman divergence is $D_\psi(x,y)=\sum_{i=1}^n x_i\ln\frac{x_i}{yi}$ between the vectors x, y (called **Kullback-Leibler** divergence). Then, we have

$$\begin{aligned}&\nabla\psi(x)=\mathrm{col}(1+\ln x_1,...,1+\ln x_n), \forall x\in\mathcal{X},\\ &\psi^*(u)=\ln\left(\sum_{i=1}^n e^{u_i}\right), \nabla\psi^*(u)=\mathrm{col}\left(\frac{e^{u_1}}{\sum_{i=1}^n e^{u_i}},...,\frac{e^{u_n}}{\sum_{i=1}^n e^{u_i}}\right),\\ &\left[\nabla^2\psi^*(u)\right]_{ij}=\frac{e^{u_i}\sum_{i=1}^n e^{u_i}-e^{u_j}e^{u_i}}{\left(\sum_{i=1}^n e^{u_i}\right)^2}, \forall\, i,j\in n.\end{aligned}$$

Furthermore, for any $i,j=1,...,n$, we have

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

Therefore, the APDMD (15) reduces to

$$\begin{cases}\ddot{x}_i(t)+\frac{\alpha+1}{t}\dot{x}_i(t)+\left(x_i(t)+\frac{t}{\alpha}\dot{x}_i(t)\right)\Big(\big(\nabla f(x(t))+\beta A^T(Ax(t)-b)\\ \qquad +A^T\left(\lambda(t)+\frac{t}{\alpha}\dot{\lambda}(t)\right)+\frac{\alpha\zeta}{t}\dot{x}(t)\big)_i \;-\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)^T\;\;(\nabla f(x(t))\\ \qquad +\beta A^T(Ax(t)-b)+A^T\left(\lambda(t)+\frac{t}{\alpha}\dot{\lambda}(t)\right)+\frac{\alpha\zeta}{t}\dot{x}(t)\Big)\Big)=0, \forall i=1,...,n,\\ \ddot{\lambda}(t)+\frac{\alpha+1+\alpha\zeta}{t}\dot{\lambda}(t)-A\left(x(t)+\frac{t}{\alpha}\dot{x}(t)\right)+b=0,\\ x(t_0)=x_0\in\mathcal{X}, \dot{x}(t_0)=\dot{x}_0, \lambda(t_0)=\lambda_0, \dot{\lambda}(t_0)=\dot{\lambda}_0,\end{cases}\tag{37}$$

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

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 $\frac{1}{2}\|x\|^2+I_{\mathcal{X}}(x)$, 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 (14) reduces to a new accelerated primal-dual projection dynamical (APDPD) approach :

$$\begin{cases} \dot{x}\left(t\right)=\frac{\alpha}{t}\left(P_{\mathcal{X}}\left(u\left(t\right)\right)-x\left(t\right)\right), \\ \dot{u}\left(t\right)=-\frac{t}{\alpha}\left(\nabla f\left(x\left(t\right)\right)+\beta A^{T}\left(Ax\left(t\right)-b\right)+A^{T}v\left(t\right)\right)-\zeta\dot{x}\left(t\right), \\ \dot{\lambda}\left(t\right)=\frac{\alpha}{t}\left(v\left(t\right)-\lambda\left(t\right)\right), \\ \dot{v}\left(t\right)=\frac{t}{\alpha}\left(AP_{\mathcal{X}}\left(u\left(t\right)\right)-b\right)-\zeta\dot{\lambda}\left(t\right), \\ x\left(t_0\right)=x_0, u\left(t_0\right)=u_0 \text{ with } x_0=P_{\mathcal{X}}\left(u_0\right)\in\mathcal{X}; \lambda\left(t_0\right)=\lambda_0, v\left(t_0\right)=0. \end{cases} \tag{38}$$

3.6 APDMD for DCCP (12) in the smooth case

In this subsection, inspired by the APDMD (14) and distributed consensus theorem, an accelerated distributed primal-dual mirror dynamical (ADPDMD) approach for DCCP (12) in the smooth case is proposed and discussed.

For DCCP (12) with smooth convex objective functions, the proposed ADPDMD with $\alpha\geqslant 2,\beta\geqslant 0,\zeta\geqslant 0$ is given by

$$\begin{cases} \dot{x}_i\left(t\right)=\frac{\alpha}{t}\left(\nabla\psi_i^*\left(u_i\left(t\right)\right)-x_i\left(t\right)\right), \\ \dot{u}_i\left(t\right)=-\frac{t}{\alpha}\Big(\nabla f_i\left(x_i\left(t\right)\right)+\beta\sum_{j=1}^{n}\mathrm{a}_{ij}\left(x_i\left(t\right)-x_j\left(t\right)\right) \\ \qquad\qquad +\sum_{j=1}^{n}\mathrm{a}_{ij}\left(\lambda_i\left(t\right)+\frac{t}{\alpha}\dot{\lambda}_i\left(t\right)-\lambda_j\left(t\right)-\frac{t}{\alpha}\dot{\lambda}_j\left(t\right)\right)\Big)-\zeta\dot{x}_i\left(t\right), \\ \dot{\lambda}_i\left(t\right)=\frac{\alpha}{t}\left(v_i\left(t\right)-\lambda_i\left(t\right)\right), \\ \dot{v}_i\left(t\right)=\frac{t}{\alpha}\sum_{j=1}^{n}\mathrm{a}_{ij}\left(\nabla\psi_i^*\left(u_i\left(t\right)\right)-\nabla\psi_j^*\left(u_j\left(t\right)\right)\right)-\zeta\dot{\lambda}_i\left(t\right), \\ x_i\left(t_0\right)=x_{i,0}, \nabla\psi_i^*\left(u_i\left(t_0\right)\right)=x_{i,0}\in\mathcal{X}_i, \\ \lambda_i\left(t_0\right)=\lambda_{i,0}, v_i\left(t_0\right)=v_{i,0}, i=1,...,n. \end{cases} \tag{39}$$

Defining $L=L_n\otimes I_m$, where L_n is the Laplacian matrix of graph $\mathcal{G}$ and let $x=\mathrm{col}\left(x_1,...,x_n\right)\in\mathbb{R}^{nm}$, $\dot{x}=\mathrm{col}\left(\dot{x}_1,...,\dot{x}_n\right)\in\mathbb{R}^{nm}$, $\lambda=\mathrm{col}\left(\lambda_1,...,\lambda_n\right)\in\mathbb{R}^{nm}$, $\dot{\lambda}=\mathrm{col}\left(\dot{\lambda}_1,...,\dot{\lambda}_n\right)\in\mathbb{R}^{nm}$, $\nabla f\left(x\right)=\mathrm{col}\left(\nabla f_1\left(x_1\right),...,\nabla f_n\left(x_n\right)\right)\in\mathbb{R}^{nm}$, $\nabla\psi^*\left(u\right)=\mathrm{col}\left(\nabla\psi_1^*\left(u_1\right),...,\nabla\psi_n^*\left(u_n\right)\right)\in\mathbb{R}^{nm}$ and $\mathcal{X}=\prod_{i=1}^{n}\mathcal{X}_i$ is the Cartesian product of set $\mathcal{X}_i$, $i=1,...,n$.

Then the compact formula of APDMD (39) is given by

$$\begin{cases} \dot{x}\left(t\right)=\frac{\alpha}{t}\left(\nabla\psi^*\left(u\left(t\right)\right)-x\left(t\right)\right), \\ \dot{u}\left(t\right)=-\frac{t}{\alpha}\left(\nabla f\left(x\left(t\right)\right)+\beta Lx\left(t\right)+Lv\left(t\right)\right)-\zeta\dot{x}\left(t\right), \\ \dot{\lambda}\left(t\right)=\frac{\alpha}{t}\left(v\left(t\right)-\lambda\left(t\right)\right), \dot{v}\left(t\right)=\frac{t}{\alpha}L\nabla\psi^*\left(u\left(t\right)\right)-\zeta\dot{\lambda}\left(t\right), \\ x\left(t_0\right)=x_0, \nabla\psi^*\left(u\left(t_0\right)\right)=x_0\in\mathcal{X}, \lambda\left(t_0\right)=\lambda_0, v\left(t_0\right)=v_0. \end{cases} \tag{40}$$

Next, we will illustrate the accelerated convergence rate of the ADPDMD (40) by the Lyapunov analysis method.

Theorem 13. *Suppose **Assumption 3.2** holds and let $\left(x\left(t\right),\lambda\left(t\right)\right)$ and $\left(x^*,\lambda^*\right)$ be a solution trajectory and an optimal solution for ADPDMD (40) and DCCP (12), respectively. Then for any $\left(x\left(t_0\right),u\left(t_0\right),\lambda\left(t_0\right),v\left(t_0\right)\right)\in\mathcal{X}\times\mathbb{R}^{nm}\times\mathbb{R}^{nm}\times\mathbb{R}^{nm}$ and let $\mathsf{L}_\beta\left(x\left(t\right),\lambda\left(t\right)\right)=f\left(x\left(t\right)\right)+\frac{\beta}{2}x\left(t\right)^TLx\left(t\right)+\lambda\left(t\right)^TLx\left(t\right)$, and consider the following candidate Lyapunov function:*

$$\mathsf{V}\left(t\right)=\mathsf{V}_1\left(t\right)+\mathsf{V}_2\left(t\right)+\mathsf{V}_3\left(t\right), \tag{41}$$

with

$$\begin{cases} \mathsf{V}_1(t) = \frac{t^2}{\alpha^2}\left(\mathsf{L}_\beta(x(t),\lambda^*) - \mathsf{L}_\beta(x^*,\lambda(t))\right) \\ \qquad\;\; = \frac{t^2}{\alpha^2}\left(f(x(t)) - f(x^*) + \frac{\beta}{2}x(t)\,Lx(t)^2 + L\lambda^*\right), \\ \mathsf{V}_2(t) = D_{\psi^*}(u(t),u^*) + \frac{\zeta}{2}\|x(t)-x^*\|^2 \\ \qquad\;\; = \psi^*(u(t)) - \psi^*(u^*) - \nabla\psi^*(u^*)^T(u(t)-u^*) + \frac{\zeta}{2}\|x(t)-x^*\|^2, \\ \mathsf{V}_3(t) = D_h(v(t),v^*) + \frac{\zeta}{2}\|\lambda(t)-\lambda^*\|^2 \\ \qquad\;\; = \frac{1}{2}\|v(t)-v^*\|^2 + \frac{\zeta}{2}\|\lambda(t)-\lambda^*\|^2, \end{cases}$$

where $u^* = \nabla\psi(x^*), v^* = \lambda^*$. *The following statements are true:*

(I) $\zeta = 0, \beta \geq 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we have*

1) $\frac{t}{\alpha}\dot{x}(t) + x(t) = \psi^*(u(t))$ *and* $\frac{t}{\alpha}\dot{\lambda}(t) + \lambda(t) = v(t)$ *of ADPDMD* (40) *are bounded;*

2) One has

$$\mathsf{L}_\beta(x(t),\lambda^*) - \mathsf{L}_\beta(x^*,\lambda(t)) \leq \frac{\alpha^2\mathsf{V}(t_0)}{t^2}; \tag{42}$$

3) if, in addition, $\beta > 0$,

$$x(t)^T Lx(t) \leq \frac{2\alpha^2\mathsf{V}(t_0)}{\beta t^2}; \tag{43a}$$

$$\int_{t_0}^{+\infty} tx(t)^T Lx(t)dt < +\infty; \tag{43b}$$

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we can get*

$$\int_{t_0}^{+\infty} t\left(\mathsf{L}_\beta(x(t),\lambda^*) - \mathsf{L}_\beta(x^*,\lambda(t))\right)dt < +\infty; \tag{44}$$

(II) $\zeta > 0,\ \beta \geqslant 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *one has*

1) $x(t)$ *and* $\lambda(t)$ *of ADPDMD* (40) *are bounded;*

$$\|\dot{x}(t)\| = O\left(\frac{1}{t}\right), \left\|\dot{\lambda}(t)\right\| = O\left(\frac{1}{t}\right); \tag{45}$$

2)

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt \leqslant \frac{\alpha}{\zeta}\mathsf{V}(t_0) < +\infty; \tag{46a}$$

$$\int_{t_0}^{+\infty} t\left\|\dot{\lambda}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\mathsf{V}(t_0) < +\infty; \tag{46b}$$

3)

$$\mathsf{L}_{\beta}\left(x\left(t\right),\lambda^{*}\right)-\mathsf{L}_{\beta}\left(x^{*},\lambda\left(t\right)\right)\leq\frac{\alpha^{2}\mathsf{V}\left(t_{0}\right)}{t^{2}};\tag{47a}$$

$$\sqrt{x\left(t\right)^{T}Lx\left(t\right)}\leqslant\frac{2\zeta\alpha^{2}\mathrm{C}_{1}}{t^{2}};\tag{47b}$$

$$\left|f\left(x\left(t\right)\right)-f\left(x^{*}\right)\right|\leqslant\frac{\alpha^{2}}{t^{2}}\left(\mathsf{V}\left(t_{0}\right)+2\zeta\left\|\lambda^{*}\right\|\mathrm{C}_{1}\right)+\frac{2\beta\zeta^{2}\alpha^{4}\mathrm{C}_{1}^{2}}{t^{4}};\tag{47c}$$

where $\mathrm{C}_1 = \sup\limits_{t\in[t_0,+\infty)}\left\{\frac{t}{\zeta\alpha}\left\|\dot{\lambda}\left(t\right)\right\|+\frac{t_0}{\zeta\alpha}\left\|\dot{\lambda}\left(t_0\right)\right\|+\frac{1+\zeta}{\zeta}\left\|\lambda\left(t\right)-\lambda\left(\ t_0\ \right)\right\|+\frac{t_0^2}{\alpha^2}\left\|Lx\left(t_0\right)\right\|\right\} < +\infty$.

(ii) $\alpha > 2$*. For any* $t \geqslant t_0 > 0, x\left(t\right)\in\mathcal{X}$*, we have*

$$\int_{t_0}^{+\infty}t\left(\mathsf{L}_{\beta}\left(x\left(t\right),\lambda^{*}\right)-\mathsf{L}_{\beta}\left(x^{*},\lambda\left(t\right)\right)\right)dt<+\infty.\tag{48}$$

Proof

Following the similar steps as the proof of **Theorem 11**, the proof can be obtained easily. Due to space limitations, we omit the proof. ∎

3.7 APDMD for DEMO (13) in smooth case

In this subsection, to address the DEMO (13), where its objective function is smooth and convex, an accelerated distributed mirror dynamical (ADMD) approach is investigated based on APDMD (14). Recalling the DEMO (13), the coupled equations $\sum_{i=1}^{n}A_ix_i = \sum_{i=1}^{n}d_i \in \mathbb{R}^m$ can be equivalently decomposed as $\bar{A}x - d + Ly = 0 \in \mathbb{R}^{nm}$ with $L = L_n \otimes I_m$ and an auxiliary variable $y = \mathrm{col}\left(y_1,...,y_n\right) \in \mathbb{R}^{nm}$ by using the properties of the Laplacian matrix of undirected $\mathcal{G}$ (i.e., $\ker\left(L_n\right) = \left\{\varsigma 1 | \varsigma \in \mathbb{R}\right\}$, $\mathrm{rang}\left(L_n\right) = \left\{\omega \in \mathbb{R}^n | \omega^T 1 = 0\right\}$). Therefore, the DEMO (13) is equivalent to:

$$\begin{aligned}&\min_{x\in\mathbb{R}^{\sum_{i=1}^{n}p_i},y\in\mathbb{R}^{nm}} f\left(x\right)=\sum_{i=1}^{n}f_i\left(x_i\right),\\&\text{s.t. } \bar{A}x-d+Ly=0, x\in\mathcal{X}=\prod_{i=1}^{n}\mathcal{X}_i,\end{aligned}\tag{49}$$

where $\bar{A} = \mathrm{bldiag}\left\{A_{p_1},A_{p_2},...,A_{p_i}\right\} \in \mathbb{R}^{nm\times\sum_{i=1}^{n}p_i}$, $d = \mathrm{col}\left(d_1,d_2,...,d_n\right) \in \mathbb{R}^{nm}$, 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 (49) with smooth convex objective functions, we propose the following accelerated distributed mirror dynamical (ADMD) approach with

$t \geq t_0 > 0$ and $\alpha \geqslant 2$ and $\zeta \geqslant 0$ as follows:

$$\begin{cases} \dot{x}(t) = \frac{\alpha}{t}\left(\nabla\psi^*(u(t)) - x(t)\right), \dot{u}(t) = -\frac{t}{\alpha}\left(\nabla f(x(t)) + \bar{A}^T v(t)\right) - \zeta\dot{x}(t), \\ \dot{\lambda}(t) = \frac{\alpha}{t}\left(v(t) - \lambda(t)\right), \dot{v}(t) = \frac{t}{\alpha}\left(\bar{A}\psi^*(u(t)) - d - \beta L\lambda(t) + Lz(t)\right) - \zeta\dot{\lambda}(t), \\ \dot{y}(t) = \frac{\alpha}{t}\left(z(t) - y(t)\right), \dot{z}(t) = -\frac{t}{\alpha}Lv(t) - \zeta\dot{y}(t), \\ x(t_0) = x_0, u(t_0) = u_0 \text{ with } \nabla\psi^*(u_0) = x_0 \in \mathcal{X}, \\ \lambda(t_0) = \lambda_0, v(t_0) = v_0, y(t_0) = y_0, z(t_0) = z_0. \end{cases} \tag{50}$$

Theorem 14. *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 (50), respectively. Then, for any $(x(t_0), u(t_0), \lambda(t_0), v(t_0), y(t_0), z(t_0)) \in \mathcal{X} \times \mathbb{R}^{\sum_{i=1}^n p_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$, and let $\mathbb{L}_\beta(x(t), \lambda(t), y(t)) = f(x(t)) - \frac{\beta}{2}\lambda(t)^T L\lambda(t) + \lambda(t)^T\left(\bar{A}x(t) - d + Ly(t)\right)$, and design a Lyapunov function as follows:*

$$\mathbb{V}(t) = \mathbb{V}_1(t) + \mathbb{V}_2(t) + \mathbb{V}_3(t) + \mathbb{V}_4(t), \tag{51}$$

with

$$\begin{cases} \mathbb{V}_1(t) = \frac{t^2}{\alpha^2}\left(\mathbb{L}_\beta(x(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda(t))\right) \\ \qquad\quad = \frac{t^2}{\alpha^2}\left(f(x(t)) - f(x^*) + (\lambda^*)^T\left(\bar{A}x(t) - d - Ly^*\right) + \frac{1}{2}\lambda(t)^T L\lambda(t)\right) \\ \mathbb{V}_2(t) = D_{\psi^*}(u(t), u^*) + \frac{\zeta}{2}\|x(t) - x^*\|^2 \\ \qquad\quad = \psi^*(u(t)) - \psi^*(u^*) - \nabla\psi^*(u^*)^T(u(t) - u^*) + \frac{\zeta}{2}\|x(t) - x^*\|^2, \\ \mathbb{V}_3(t) = D_h(v(t), v^*) + \frac{\zeta}{2}\|\lambda(t) - \lambda^*\|^2 = \frac{1}{2}\|v(t) - v^*\|^2 + \frac{\zeta}{2}\|\lambda(t) - \lambda^*\|^2, \\ \mathbb{V}_4(t) = D_h(z(t), z^*) + \frac{\zeta}{2}\|y(t) - y^*\|^2 = \frac{1}{2}\|z(t) - z^*\|^2 + \frac{\zeta}{2}\|y(t) - y^*\|^2, \end{cases}$$

where $u^* = \nabla\psi(x^*)$, $v^* = \lambda^*$ and $z^* = y^*$. We have the following statements:

(I) $\zeta = 0$, $\beta \geq 0$.

(i) $\alpha \geqslant 2$. For any $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, we have

1) $\frac{t}{\alpha}\dot{x}(t) + x(t) = \psi^*(u(t))$, $\frac{t}{\alpha}\dot{\lambda}(t) + \lambda(t) = v(t)$ and $\frac{t}{\alpha}\dot{y}(t) + y(t) = z(t)$ of ADMD (50) are bounded;

2) One has

$$\mathbb{L}_\beta(x(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda(t)) \leq \frac{\alpha^2\mathbb{V}(t_0)}{t^2}; \tag{52}$$

3) if, in addition, $\beta > 0$,

$$\lambda(t)^T L\lambda(t) \leq \frac{2\alpha^2\mathbb{V}(t_0)}{\beta t^2}; \tag{53a}$$

$$\int_{t_0}^{+\infty} t\lambda(t)^T L\lambda(t)dt < +\infty; \tag{53b}$$

(ii) $\alpha > 2$. For any $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, we can get

$$\int_{t_0}^{+\infty} t\left(\mathbb{L}_\beta(x(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda(t))\right)dt < +\infty; \tag{54}$$

(II) $\zeta > 0$, $\beta \geq 0$.

(i) $\alpha \geqslant 2$. For any $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, one has

1) $x(t)$, $y(t)$ and $\lambda(t)$ of ADMD (50) are bounded;

$$\|\dot{x}(t)\| = O\left(\frac{1}{t}\right), \|\dot{y}(t)\| = O\left(\frac{1}{t}\right), \left\|\dot{\lambda}(t)\right\| = O\left(\frac{1}{t}\right); \tag{55}$$

2)

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt \leqslant \frac{\alpha}{\zeta}\mathbb{V}(t_0) < +\infty; \tag{56a}$$

$$\int_{t_0}^{+\infty} t\|\dot{y}(t)\|^2 dt \leqslant \frac{\alpha}{\zeta}\mathbb{V}(t_0) < +\infty; \tag{56b}$$

$$\int_{t_0}^{+\infty} t\left\|\dot{\lambda}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\mathbb{V}(t_0) < +\infty; \tag{56c}$$

3)

$$\mathbb{L}_\beta(x(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda(t)) \leq \frac{\alpha^2\mathbb{V}(t_0)}{t^2}; \tag{57}$$

(ii) $\alpha > 2$. For any $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, we have

$$\int_{t_0}^{+\infty} t(\mathbb{L}_\beta(x(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda(t)))dt < +\infty. \tag{58}$$

Proof Using the same proof steps as **Theorem 11** with $\bar{A}x^* = d + Ly^*$, $v^* = \lambda^*$, $y^* = z^*$, $\nabla\psi^*(u^*) = x^*$, and $L\lambda^* = 0$, we can get

$$\begin{aligned}
\dot{\mathbb{V}}(t) =& \dot{\mathbb{V}}_1(t) + \dot{\mathbb{V}}_2(t) + \dot{\mathbb{V}}_3(t) + \dot{\mathbb{V}}_4(t) \\
=& \frac{2t}{\alpha^2}\left(f(x(t)) - f(x^*) + (\lambda^*)^T\left(\bar{A}x(t) - d - Ly^*\right) + \frac{\beta}{2}\lambda(t)^T L\lambda(t)\right) \\
&+ \frac{t}{\alpha}\left(\nabla f(x(t)) + \bar{A}^T\lambda^*\right)^T(x^* - x(t)) \\
&+ \frac{\beta t}{\alpha}(L\lambda(t))^T(v(t) - v^*) + \frac{\beta t}{\alpha}(L\lambda(t))^T(v^* - \lambda(t)) \\
&- \frac{t}{\alpha}\left(\bar{A}\nabla\psi^*(u(t)) - d - Ly^*\right)^T(v(t) - v^*) + \frac{t}{\alpha}(y^*)^T Lv(t) \\
&+ \frac{t}{\alpha}(v(t) - v^*)^T\left(\bar{A}\psi^*(u(t)) - d - \beta L\lambda(t)\right) \\
&+ \frac{t}{\alpha}(v(t) - v^*)^T Lz(t) - \frac{t}{\alpha}(z(t) - z^*)^T Lv(t) - \frac{t}{\alpha}(z^*)^T Lv(t) \\
&- \frac{t\zeta}{\alpha}\left\|\dot{\lambda}(t)\right\|^2 - \frac{t\zeta}{\alpha}\|\dot{y}(t)\|^2 - \frac{t\zeta}{\alpha}\|\dot{x}(t)\|^2 \\
\leqslant& -\frac{(\alpha-2)t}{\alpha^2}(\mathbb{L}_\beta(x(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda(t))) \\
&- \frac{\beta t}{2\alpha}\lambda(t)^T L\lambda(t) - \frac{t\zeta}{\alpha}\left\|\dot{\lambda}(t)\right\|^2 - \frac{t\zeta}{\alpha}\|\dot{y}(t)\|^2 - \frac{t\zeta}{\alpha}\|\dot{x}(t)\|^2 \\
\leqslant& 0.
\end{aligned} \tag{59}$$

The rest results of proof follows from **Theorem 11**, which we can easily prove. Due to space limitations, we omit the proof. ■

4. Optimization approaches for problem (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 (14) to this nonsmooth case, we will consider a corresponding smoothing accelerated primal-dual mirror dynamical (SAPDMD) approaches based on smoothing approximation in (2.6) as follows:

$$\begin{cases} \dot{x}^{\mu}(t) = \frac{\alpha}{t}\left(\nabla\psi^{*}\left(u^{\mu}(t)\right) - x^{\mu}(t)\right), \\ \dot{u}^{\mu}(t) = -\frac{t}{\alpha}\left(\nabla_{x}\hat{f}\left(x^{\mu}(t), \mu(t)\right) + \beta A^{T}\left(Ax^{\mu}(t) - b\right) + A^{T}v^{\mu}(t)\right) - \zeta\dot{x}^{\mu}(t), \\ \dot{\lambda}^{\mu}(t) = \frac{\alpha}{t}\left(v^{\mu}(t) - \lambda^{\mu}(t)\right), \\ \dot{v}^{\mu}(t) = \frac{t}{\alpha}\left(A\nabla\psi^{*}\left(u^{\mu}(t)\right) - b(t)\right) - \zeta\dot{v}^{\mu}(t), \\ x^{\mu}(t_0) = x_0^{\mu}, u^{\mu}(t_0) = u_0^{\mu}, \nabla\psi^{*}\left(u_0^{\mu}\right) = x_0^{\mu}, \\ \lambda^{\mu}(t_0) = \lambda_0^{\mu}, v^{\mu}(t_0) = v_0^{\mu}, \mu(t_0) = \mu_0, \end{cases} \tag{60}$$

where $t \geq t_0 > 0$, $\beta \geq 0$, $\zeta \geq 0$, $\alpha \geq 2$ and $\mu(t) \leq \frac{\mu_0}{t^{2\alpha}}$. The superscript μ 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 (14), ADPDMD (40) and ADMD (50). The $\nabla_x\hat{f}\left(x^{\mu}(t), \mu(t)\right)$ is the gradient of smoothing function of $\hat{f}\left(x^{\mu}(t), \mu(t)\right)$ with respect of $x^{\mu}(t)$ for fixed $\mu(t)$. The SAPDMD (60) is illustrated in Figure 4: (left).

Similar to APDMD (14), the SAPDMD (60) can be equivalent to the following smoothing second-order dynamical approach:

$$\begin{cases} \ddot{x}^{\mu}(t) + \frac{\alpha+1}{t}\dot{x}^{\mu}(t) + \nabla^2\psi^{*}\left(\nabla\psi\left(x^{\mu}(t) + \frac{t}{\alpha}\dot{x}^{\mu}(t)\right)\right)\left(\nabla_x\hat{f}\left(x^{\mu}(t), \mu(t)\right)\right. \\ \qquad \left. + \beta A^{T}\left(Ax^{\mu}(t) - b\right) + A^{T}\left(\lambda^{\mu}(t) + \frac{t}{\alpha}\dot{\lambda}^{\mu}(t)\right)\right) \\ \qquad + \frac{\alpha\zeta}{t}\nabla^2\psi^{*}\left(\nabla\psi\left(x^{\mu}(t) + \frac{t}{\alpha}\dot{x}^{\mu}(t)\right)\right)\dot{x}^{\mu}(t) = 0, \\ \ddot{\lambda}^{\mu}(t) + \frac{\alpha+1+\alpha\zeta}{t}\dot{\lambda}^{\mu}(t) - A\left(x^{\mu}(t) + \frac{\alpha}{t}\dot{x}^{\mu}(t)\right) + b = 0, \\ x^{\mu}(t_0) = x_0^{\mu} \in \mathcal{X}, \dot{x}^{\mu}(t_0) = \dot{x}_0^{\mu}, \lambda^{\mu}(t_0) = \lambda_0^{\mu}, \dot{\lambda}^{\mu}(t_0) = \dot{\lambda}_0^{\mu}. \end{cases} \tag{61}$$

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 in (Cabot and Paoli (2007); Vassilis et al. (2018); He et al. (2017)), Moreau-Yosida regularization method in (Balavoine et al. (2013); Attouch et al. (2022)) and directional derivative method in (Su et al. (2016); Fazlyab et al. (2017)).

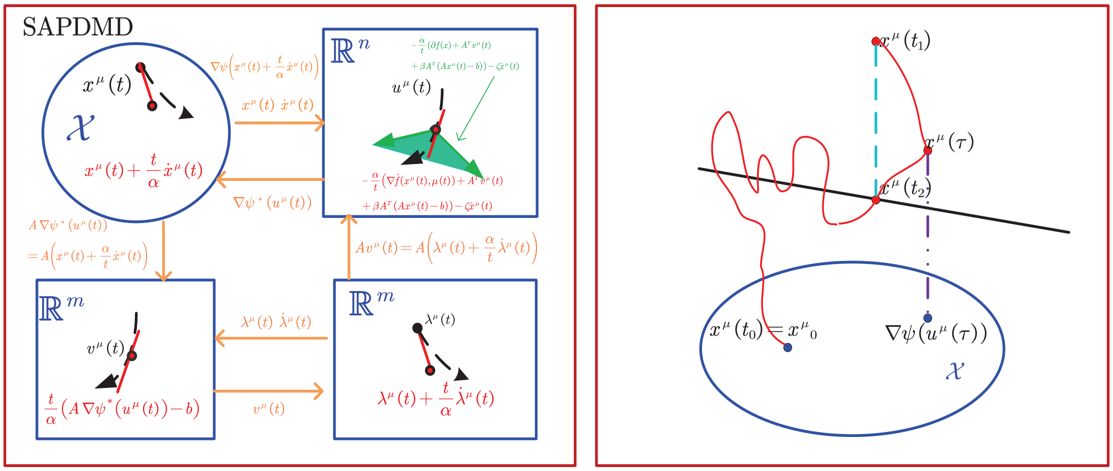

Figure 4: (left) The demonstration of SAPDMD (60). (right) The demonstration of the proof of feasibility of SAPDMD (60)

.

Compared with the dynamical approaches based on techniques mentioned above for solving nonsmooth optimization problems, our proposed SAPDMD (60) based on smoothing approximation technology in **Section** 2.6 has the following differences and preponderances.

- The solutions may be different. The SAPDMD (60) has a strong global solution, which is similar to that in (9). In (Vassilis et al. (2018); Cabot and Paoli (2007)), some inertial differential inclusion dynamical approaches are investigated, but they have shock solutions. The dynamical differential inclusion approaches proposed in (He et al. (2017); Zeng et al. (2018)) have a Carathéodory's solution or Filippov's solution.

- The SAPDMD (60) can achieve the optimal solution of problem (1) in the nonsmooth case when $\mu \to 0$, which is different from the works in (Nesterov (2005)) and (Allen-Zhu and Hazan (2016)). In the famous work (Nesterov (2005)), 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 ε-solution, i.e., $f\left(x^{\varepsilon}\right) - \min f\left(x\right) \leq \varepsilon$, where ε is a small constant, but it's not equal to 0.

- The proposed SAPDMD (60) does not require solving some subproblems. However, the accelerated dynamic approaches in (Balavoine et al. (2013)) and (Attouch et al. (2022)) need to use the Moreau-Yosida approximation $f_{\rho}\left(x\right) = \min\limits_{x \in \mathcal{X}} \left\{ f\left(\xi\right) + \frac{1}{2\rho} \left\|x - \xi\right\|^2 \right\}$, the accelerated dynamical aprroaches in (Su et al. (2016)) and (Fazlyab et al. (2017)) need to utilize $d\left(x; \dot{x}\right) = \arg\max\limits_{f \in \partial f(x)} f^T \dot{x}$, 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 (60) can be easily guaranteed by the Cauchy-Lipschitz-Picard theorem. However, the uniqueness of the solutions in directional derivative methods in (Su et al. (2016)) and (Fazlyab et al. (2017)), and differential inclusion methods in (He et al. (2017)) 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 of $x^{\mu}(t)$ for the SAPDMD (60) by the same way as in **Section** 3.3.

Lemma 15. *For any $(x^{\mu}(t_0), u^{\mu}(t_0), \lambda^{\mu}(t_0), \upsilon^{\mu}(t_0)) \in \mathcal{X}\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$ with $\nabla\psi^*\left(u^{\mu}_{t_0}\right) = x^{\mu}_{t_0}$, 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** 8 by the reductio and separation hyperplane theorem, we omit it here due to the limitation of space.* ■

Now let us turn to the existence of strong global solution for the problem (1) with nonsmooth objective functions and we will once again take into account solutions (similar to **Definition** 9 to this problem.

Theorem 16. *For any initial values $(x^{\mu}(t_0), u^{\mu}(t_0), \lambda^{\mu}(t_0), \upsilon^{\mu}(t_0)) \in \mathcal{X}\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$, there exists a unique strong global solution of SAPDMD* (60)*.*

Proof Let $Y^{\mu}(t) = (x^{\mu}(t), u^{\mu}(t), \lambda^{\mu}(t), \upsilon^{\mu}(t))$, then the SAPDMD (60) is equivalent to

$$\begin{cases} \dot{Y}^{\mu}(t) = H(t, Y^{\mu}(t)) \\ Y^{\mu}(t_0) = (x^{\mu}(t_0), u^{\mu}(t_0), \lambda^{\mu}(t_0), \upsilon^{\mu}(t_0)), \end{cases} \tag{62}$$

where

$$\begin{aligned} H(t, Y^{\mu}) = \Big(& \frac{\alpha}{t}\left(\nabla\psi^*(u^{\mu}(t)) - x^{\mu}(t)\right),\ -\frac{t}{\alpha}\Big(\nabla_x \hat{f}(x^{\mu}(t), \mu(t)) \\ & - A^T\upsilon^{\mu}(t)\Big) - \zeta\dot{x}^{\mu}(t),\ \frac{\alpha}{t}\left(\upsilon^{\mu}(t) - \lambda^{\mu}(t)\right),\ \frac{t}{\alpha}\left(A\nabla\psi^*(u^{\mu}(t)) - b\right) - \zeta\dot{\lambda}^{\mu}(t)\Big). \end{aligned} \tag{63}$$

To prove the existence and uniqueness of the strong global solution $Y^{\mu}(t)$ generated by SAPDMD (60) by the Cauchy-Lipschitz-Picard theorem in (Bolte (2003)), the following conditions need to be satisfied:

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

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

The proof of (I). Let $t \in [t_0, +\infty)$ be fixed and use the Lipschitz continuous of $\nabla\psi^*$, $\nabla_x f(x,\mu)$. Then, for any Y^μ, $\hat{Y}^\mu$, we have

$$\left\|F\left(t, Y^\mu(t)\right) - F\left(t, \hat{Y}^\mu(t)\right)\right\| \leq \left(\frac{t^2}{\alpha^2}\left(4\beta^2\delta_{\max}\left(A^TA\right)^2 + \frac{4\ell^2}{\mu^2(t)} + \delta_{\max}\left(A^TA\right)\mathfrak{l}_{\psi^*}^2 + 2\delta_{\max}\left(A^TA\right)\right) + \frac{\alpha^2\left(6+2\zeta+(2+\zeta)\mathfrak{l}_{\psi^*}^2\right)}{t^2}\right)^{\frac{1}{2}}\left\|Y^\mu(t) - \hat{Y}^\mu(t)\right\|.$$

By using the notation $\mathfrak{l}(t) = \left(\frac{t^2\left(4\beta^2\delta_{\max}\left(A^TA\right)^2 + \frac{4\ell^2}{\mu^2(t)} + \delta_{\max}\left(A^TA\right)\mathfrak{l}_{\psi^*}^2 + 2\delta_{\max}\left(A^TA\right)\right)}{\alpha^2} + \frac{\alpha^2\left(6+2\zeta+(2+\zeta)\mathfrak{l}_{\psi^*}^2\right)}{t^2}\right)^{\frac{1}{2}}$, one has $\left\|H\left(t, Y^\mu(t)\right) - H\left(t, \hat{Y}^u(t)\right)\right\| \leq \mathfrak{l}(t)\left\|Y^\mu(t) - \hat{Y}^\mu(t)\right\|$.

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

The proof of (II). Let $Y \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ and $t_0 < \mathcal{T} < +\infty$, it holds that

$$\begin{aligned}
&\int_{t_0}^{\mathcal{T}} \|H\left(t, Y^\mu(t)\right)\| dt \\
\leq &\sqrt{\|\nabla\psi^*\left(u(t)\right)\|^2 + \|x(t)\|^2 + \|\lambda(t)\|^2 + \|v(t)\|^2 + \|x^*\|^2 + \left\|\nabla_x \hat{f}\left(x^\mu(t), \mu(t)\right)\right\|^2} \\
&\times \left(\frac{(8+4\zeta)\alpha^2}{t^2} + \frac{2t^2}{\alpha^2}\left(4\delta_{\max}\left(A^TA\right) + 4\beta^2\delta_{\max}\left(A^TA\right)^2 + 2\right)\right),
\end{aligned}$$

and the conclusion holds by employing the continuity of the function

$$t \to \left(\frac{(8+4\zeta)\alpha^2}{t^2} + \frac{2t^2}{\alpha^2}\left(4\delta_{\max}\left(A^TA\right) + 4\beta^2\delta_{\max}\left(A^TA\right)^2 + 2\right)\right)^{\frac{1}{2}} \text{ on } [t_0, \mathcal{T}].$$

The existence and uniqueness of $Y^\mu(t)$ to the dynamical system (63) can be guaranteed by the Cauchy-Lipschitz-Picard theorem, consequently, the existence and uniqueness of the trajectories of SAPDMD (60) also hold. ∎

4.3 The accelerated convergence of the SAPDMD

In this subsection, we will illustrate the accelerated convergence properties of the SAPDMD (60) 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., $\nabla_\mu \hat{f}(x(t), \mu(t))$ needs to be considered. The third factor is boundedness of gradient with respect to $\mu(t)$. In turn, all the results and estimation we have presented in the previous sections can be transferred to this more general nonsmooth context.

Theorem 17. *Suppose that **Assumption 3.1** holds and the objective function is nonsmooth convex. Let $(x(t), \lambda(t))$ and (x^*, λ^*) be a solution trajectory and an optimal solution of SAPDMD (60), respectively. Defining a Lyapunov function as $\hat{\mathcal{L}}_\beta(x^\mu(t), \lambda^\mu(t)) = \hat{f}(x^\mu(t), \mu(t)) + \lambda^\mu(t)^T(Ax^\mu(t) - b) + \frac{\beta}{2}\|Ax^\mu(t) - b\|^2$, $\mathcal{L}_\beta(x^\mu(t), \lambda^\mu(t)) = f(x^\mu(t)) + \lambda^\mu(t)^T(Ax^\mu(t) - b) + \frac{\beta}{2}\|Ax^\mu(t) - b\|^2$ with $\beta \geq 0$, then, for any $(x^\mu(t), \lambda^\mu(t)) \in \mathcal{X} \times \mathbb{R}^m$, and consider a well-designed smooth Lyapunov function*

$$\hat{\mathcal{V}}(t) = \hat{\mathcal{V}}_1(t) + \hat{\mathcal{V}}_2(t) + \hat{\mathcal{V}}_3(t), \tag{64}$$

with

$$\begin{cases} \hat{\mathcal{V}}_1(t) = \frac{t^2}{\alpha^2}\Big(\hat{\mathcal{L}}_\beta(x^\mu(t), \lambda^*) - \hat{\mathcal{L}}_\beta(x^*, \lambda^\mu(t)) + 4\kappa_{\hat{f}}\mu(t)\Big) \\ \qquad\quad = \frac{t^2}{\alpha^2}\Big(\hat{f}(x^\mu(t), \mu(t)) - f(x^*, \mu(t)) + 4\kappa_{\hat{f}}\mu(t) \\ \qquad\qquad + \frac{\beta}{2}\|Ax^\mu(t) - b\|^2 + (\lambda^*)^T A(x^\mu(t) - x^*)\Big) \\ \hat{\mathcal{V}}_2(t) = D_{\psi^*}(u^\mu(t), u^*) + \frac{\zeta}{2}\|x^\mu(t) - x^*\|^2 \\ \qquad\quad = \psi^*(u^\mu(t)) - \psi^*(u^*) - \nabla\psi^*(u^*)^T(u^\mu(t) - u^*) + \frac{\zeta}{2}\|x^\mu(t) - x^*\|^2, \\ \hat{\mathcal{V}}_3(t) = D_h(v^\mu(t), v^*) + \frac{\zeta}{2}\|\lambda^\mu(t) - \lambda^*\|^2 = \frac{1}{2}\|v^\mu(t) - v^*\|^2 + \frac{\zeta}{2}\|\lambda^\mu(t) - \lambda^*\|^2. \end{cases}$$

where $u^* = \nabla\psi(x^*)$, $v^* = \lambda^*$. The following statements are true:

(I) $\zeta = 0, \beta \geq 0$.

(i) $\alpha \geqslant 2$. For any $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, we have

1) $\frac{t}{\alpha}\dot{x}^\mu(t) + x^\mu(t) = \psi^*(u^\mu(t))$ and $\frac{t}{\alpha}\dot{\lambda}^\mu(t) + \lambda^\mu(t) = v^\mu(t)$ of SAPDMD (60) are bounded;

2) One has

$$|\mathcal{L}_\beta(x^\mu(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda^\mu(t))| \leqslant \frac{\alpha^2\hat{\mathcal{V}}(t_0)}{t^2} + \frac{2\kappa_{\hat{f}}\mu_0}{t^{2\alpha}}; \tag{65}$$

3) if, in addition, $\beta > 0$,

$$\|Ax^\mu(t) - b\|^2 \leq \frac{2\alpha^2\hat{\mathcal{V}}(t_0)}{\beta t^2}; \tag{66a}$$

$$\int_{t_0}^{+\infty} t\,\|Ax^\mu(t) - b\|^2 dt < +\infty; \tag{66b}$$

(ii) $\alpha > 2$. For any $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, we have

$$\int_{t_0}^{+\infty} t\Big(\mathcal{L}_\beta(x^\mu(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda^\mu(t)) + 2\kappa_{\hat{f}}\mu(t)\Big)dt < +\infty; \tag{67}$$

(II) $\zeta > 0, \beta \geqslant 0$.

(i) $\alpha \geqslant 2$. For any $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, one has

1) $x^\mu(t)$ and $\lambda^\mu(t)$ are bounded;

$$\|\dot{x}^\mu(t)\| = O\left(\frac{1}{t}\right), \left\|\dot{\lambda}^\mu(t)\right\| = O\left(\frac{1}{t}\right); \tag{68}$$

2)

$$\int_{t_0}^{+\infty} t \left\| \dot{x^{\mu}} (t) \right\|^2 dt \leqslant \frac{\alpha}{\zeta} \hat{\mathcal{V}} (t_0) < +\infty; \tag{69a}$$

$$\int_{t_0}^{+\infty} t \left\| \dot{\lambda^{\mu}} (t) \right\|^2 dt \leqslant \frac{\alpha}{\zeta} \hat{\mathcal{V}} (t_0) < +\infty; \tag{69b}$$

3)

$$\left| \mathcal{L}_{\beta} (x^{\mu} (t), \lambda^*) - \mathcal{L}_{\beta} (x^*, \lambda^{\mu} (t)) \right| \leqslant \frac{\alpha^2 \hat{\mathcal{V}} (t_0)}{t^2} + \frac{2\kappa_{\hat{f}} \mu_0}{t^{2\alpha}}; \tag{70a}$$

$$\left\| A x^{\mu} (t) - b \right\| \leqslant \frac{2\zeta \alpha^2 \mathcal{C}_1}{t^2}; \tag{70b}$$

$$\left| f (x^{\mu} (t)) - f (x^*) \right| \leqslant \frac{\alpha^2}{t^2} \left(\hat{\mathcal{V}} (t_0) + 2\zeta \left\| \lambda^* \right\| \mathcal{C}_1 \right) + \frac{2\beta \zeta^2 \alpha^4 \mathcal{C}_1^2}{t^4}; \tag{70c}$$

where $\mathcal{C}_1 = \sup\limits_{t \in [t_0, +\infty)} \left\{ \frac{t}{\zeta\alpha} \left\| \dot{\lambda}^{\mu} (t) \right\| + \frac{t_0}{\zeta\alpha} \left\| \dot{\lambda}^{\mu} (t_0) \right\| + \frac{1+\zeta}{\zeta} \left\| \lambda^{\mu} (t) - \lambda^{\mu} (t_0) \right\| + \frac{t_0^2}{\alpha^2} \left\| A x^{\mu} (t_0) - b \right\| \right\} < +\infty$.

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we have*

$$\int_{t_0}^{+\infty} t \Big(\mathcal{L}_{\beta} (x^{\mu} (t), \lambda^*) - \mathcal{L}_{\beta} (x^*, \lambda^{\mu} (t)) + 2\kappa_{\hat{f}} \mu (t) \Big) dt < +\infty. \tag{71}$$

Proof

Note that the function $\hat{\mathcal{V}}_1 (t)$ is nonnegative for any $t \geq t_0 > 0$ since

$$\begin{aligned} &\hat{\hat{\mathcal{L}}}_{\beta} (x^{\mu} (t), \lambda^*) - \hat{\mathcal{L}}_{\beta} (x^*, \lambda^{\mu} (t)) + 4\kappa_{\hat{f}} \mu (t) \\ \geq& f (x^{\mu} (t)) - f (x^*) + (\lambda^*)^T A (x^{\mu} (t) - x^*) + \frac{\beta}{2} \left\| A x^{\mu} (t) - b \right\|^2 + 2\kappa_{\hat{f}} \mu (t) \\ =& \mathcal{L}_{\beta} (x^{\mu} (t), \lambda^*) - \mathcal{L}_{\beta} (x^*, \lambda^{\mu} (t)) + 2\kappa_{\hat{f}} \mu (t) \geq 0, \end{aligned} \tag{72}$$

where the first equality holds from $\hat{f} (x^{\mu} (t), \mu (t)) - f (x^{\mu} (t)) + \kappa_{\hat{f}} \mu (t) \geq 0$, $f (x^*) - f (x^*, \mu (t)) + \kappa_{\hat{f}} \mu (t) \geq 0$ in **Definition 6** *(iv)*, and the second inequality is satisfied due to inequalities (9). Thus the $\hat{\mathcal{V}}_1 (t) \geqslant 0$ holds, i.e., $\hat{\mathcal{V}}_1 (t)$ is nonnegative.

When $\zeta = 0$. Since $\psi^* (\cdot)$ is convex, it gives $D_{\psi^*} (u^{\mu} (t), u^*) \geq 0$, and $h (\cdot) = \frac{1}{2} \left\| \cdot \right\|^2$ is strongly convex, we can obtain that $D_h (v^{\mu} (t), \lambda^*) \geq 0$ and $\left\| v^{\mu} (t) \right\| \to +\infty$, $\hat{\mathcal{V}}_2 (t) \to +\infty$, i.e., $\hat{\mathcal{V}}_2 (t)$ is radially unbounded of variable $v^{\mu} (t)$. Thus, $\hat{\mathcal{V}} (t)$ is nonnegative and radially unbounded of variable $v^{\mu} (t)$.

When $\zeta > 0$. According to $\frac{\zeta}{2} \left\| x^{\mu} (t) - x^* \right\|^2$, one has $\left\| x^{\mu} (t) \right\| \to +\infty, \hat{\mathcal{V}}_2 (t) \to +\infty$ which means $\hat{\mathcal{V}}_2 (t)$ is nonnegative and radially unbounded of $x^{\mu} (t)$. In addition, for the $\hat{\mathcal{V}}_3 (t)$, since $D_h (v^{\mu} (t), v^*) = \frac{1}{2} \left\| v^{\mu} (t) - v^* \right\|^2$ and $\frac{\zeta}{2} \left\| \lambda^{\mu} (t) - \lambda^* \right\|^2 \geqslant 0$ are strongly convex functions of variables $v^{\mu} (t)$ and $\lambda^{\mu} (t)$, it implies $\left\| v^{\mu} (t) \right\| \to +\infty$ or $\left\| \lambda^{\mu} (t) \right\| \to +\infty$, $\hat{\mathcal{V}}_3 (t) \to +\infty$. Thus, $\hat{\mathcal{V}}_3 (t)$ is nonnegative and radially unbounded of variables $v^{\mu} (t)$ and $\lambda^{\mu} (t)$. Thus, $\hat{\mathcal{V}} (t)$ is nonnegative and radially unbounded of variables $x^{\mu} (t)$, $v^{\mu} (t)$ and $\lambda^{\mu} (t)$.

The derivatives of $\hat{\mathcal{V}}_1(t)$, $\hat{\mathcal{V}}_2(t)$ and $\hat{\mathcal{V}}_3(t)$ along the trajectory of SAPDMD (60) with $\nabla\psi^*(u^*) = x^*$ and $Ax^* = b$ satisfy

$$
\begin{aligned}
\dot{\hat{\mathcal{V}}}_1(t) =& \frac{2t}{\alpha^2}\Big(\hat{f}(x^\mu(t),\mu(t)) - f(x^*,\mu(t)) + (\lambda^*)^T A(x^\mu(t)-x^*) \\
&+4\kappa_{\hat{f}}\mu(t) + \frac{\beta}{2}\|Ax^\mu(t)-b\|^2\Big) + \frac{t}{\alpha}(\nabla\psi^*(u^\mu(t)) - \nabla\psi^*(u^*))^T \\
&\times\Big(\nabla_x\hat{f}(x^\mu(t),\mu(t)) + \beta A^T(Ax^\mu(t)-b) + A^T\lambda^*\Big) \\
&+\frac{t}{\alpha}\Big(\nabla_x\hat{f}(x^\mu(t),\mu(t)) + \beta A^T(Ax^\mu(t)-b) + A^T\lambda^*\Big)^T \\
&\times(\nabla\psi^*(u^*) - x^\mu(t)) + \frac{2t^2}{\alpha^2}\kappa_{\hat{f}}\dot{\mu}(t) \\
&+\frac{t^2}{\alpha^2}\Big(\nabla_\mu\hat{f}(x^\mu(t),\mu(t)) + \nabla_\mu\hat{f}(x^*,\mu(t)) + 2\kappa_{\hat{f}}\Big)\dot{\mu}(t),
\end{aligned}
\tag{73}
$$

$$
\begin{aligned}
\dot{\hat{\mathcal{V}}}_2(t) =& -\zeta(\nabla\psi^*(u^\mu(t)) - \nabla\psi^*(u^*))^T\dot{x}^\mu(t) \\
&-\frac{t}{\alpha}\Big(\nabla_x\hat{f}(x^\mu(t),\mu(t)) + \beta A^T A(x^\mu(t)-x^*) + A^T v^\mu(t)\Big)^T \\
&\times(\nabla\psi^*(u^\mu(t)) - \nabla\psi^*(u^*)) + \zeta(x^\mu(t)-x^*)^T\dot{x}^\mu(t) \\
=& -\frac{t}{\alpha}(\nabla\psi^*(u^\mu(t)) - \nabla\psi^*(u^*))^T\Big(A^T v^\mu(t) - A^T\lambda^*\Big) \\
&-\frac{t}{\alpha}\Big(\nabla_x\hat{f}(x^\mu(t),\mu(t)) + \beta A^T A(x^\mu(t)-x^*) + A^T\lambda^*\Big)^T \\
&\times(\nabla\psi^*(u^\mu(t)) - \nabla\psi^*(u^*)) - \frac{\zeta t}{\alpha}\|\dot{x}^\mu(t)\|^2,
\end{aligned}
\tag{74}
$$

$$
\begin{aligned}
\dot{\hat{\mathcal{V}}}_3(t) =& \frac{t}{\alpha}(v^\mu(t)-\lambda^*)^T A(\nabla\psi^*(u^\mu(t)) - x^*) \\
&-(v^\mu(t)-\lambda^*)^T\dot{\lambda}^\mu(t) + \zeta(\lambda^\mu(t)-\lambda^*)^T\dot{\lambda}^\mu(t) \\
=& \frac{t}{\alpha}(v^\mu(t)-\lambda^*)^T A(\nabla\psi^*(u^\mu(t)) - x^*) - \frac{\zeta t}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2
\end{aligned}
\tag{75}
$$

Combining (73), (74), (75) with $\nabla\psi^*(u^*) = x^*, Ax^* = b, v^* = \lambda^*$ and rearranging them yields

$$
\begin{aligned}
\dot{\hat{\mathcal{V}}}(t) =& \dot{\hat{\mathcal{V}}}_1(t) + \dot{\hat{\mathcal{V}}}_2(t) + \dot{\hat{\mathcal{V}}}_3(t) \\
\le& \frac{2t}{\alpha^2}\Big(\hat{f}(x^\mu(t),\mu(t)) - \hat{f}(x^*,\mu(t)) + \frac{\beta}{2}\|Ax^\mu(t)-b\|^2 + 4\kappa_{\hat{f}}\mu(t) \\
&+(\lambda^*)^T A(x^\mu(t)-x^*)\Big) - \frac{t}{\alpha}\Big(\hat{f}(x^\mu(t),\mu(t)) - \hat{f}(x^*,\mu(t)) \\
&+\beta\|Ax^\mu(t)-b\|^2 + (\lambda^*)^T A(x^\mu(t)-x^*)\Big) + \frac{2t^2}{\alpha^2}\kappa_{\hat{f}}\dot{\mu}(t) \\
&-\frac{\zeta t}{\alpha}\|\dot{x}^\mu(t)\|^2 - \frac{\zeta t}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
\leq & \frac{(2-\alpha)\, t}{\alpha^2}\left(\hat{f}\left(x^\mu(t), \mu(t)\right)-\hat{f}\left(x^*, \mu(t)\right)^2+4 \kappa_{\hat{f}} \mu(t)\right. \\
& \left.+\left(\lambda^*\right)^T A\left(x^\mu(t)-x^*\right)+\frac{\beta}{2}\left\|A x^\mu(t)-b\right\|\right) \\
& -\frac{t \beta}{2 \alpha}\left\|A x^\mu(t)-b\right\|^2+\frac{4 t}{\alpha} \kappa_{\hat{f}} \mu(t)+\frac{2 t^2}{\alpha^2} \kappa_{\hat{f}} \dot{\mu}(t) \\
& -\frac{\zeta t}{\alpha}\left\|\dot{x}^\mu(t)\right\|^2-\frac{\zeta t}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2 \\
\leq & -\frac{(\alpha-2)\, t}{\alpha^2}\left(\hat{\mathcal{L}}_\beta\left(x^\mu(t), \lambda^*\right)-\hat{\mathcal{L}}_\beta\left(x^*, \lambda^\mu(t)\right)+4 \kappa_{\hat{f}} \mu(t)\right) \\
& -\frac{t \beta}{2 \alpha}\left\|A x^\mu(t)-b\right\|^2 \leq 0-\frac{\zeta t}{\alpha}\left\|\dot{x}^\mu(t)\right\|^2-\frac{\zeta t}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2 \\
\leq & -\frac{(\alpha-2)\, t}{\alpha^2}\left(\mathcal{L}_\beta\left(x^\mu(t), \lambda^*\right)-\mathcal{L}_\beta\left(x^*, \lambda^\mu(t)\right)+2 \kappa_{\hat{f}} \mu(t)\right) \\
& -\frac{t \beta}{2 \alpha}\left\|A x^\mu(t)-b\right\|^2 \leq 0-\frac{\zeta t}{\alpha}\left\|\dot{x}^\mu(t)\right\|^2-\frac{\zeta t}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2,
\end{aligned}
\tag{76}
$$

where the first inequality holds since $\nabla_\mu \hat{f}\left(x^\mu(t), \mu(t)\right)+\nabla_\mu \hat{f}\left(x^*, \mu(t)\right)+2 \kappa_{\hat{f}} \geq 0$ and $\dot{\mu}(t) \leq 0$, the second inequality is satisfied because of the convexity of $\hat{f}\left(x^\mu(t), \mu(t)\right)$ of $x^\mu(t) \in \mathcal{X}$ with any fixed $\mu(t) \in\left(0, \mu_0\right]$, the third inequality holds from $\mu(t) \leq \frac{\mu_0}{t^{2 \alpha}}$ (i.e., $\dot{\mu}(t) \leq-\mu_0(2 \alpha)\, t^{(-2 \alpha-1)}$ and $\frac{t^2}{\alpha^2} \kappa_{\hat{f}} \geq 0$ imply $\frac{2 t}{\alpha} \kappa_{\hat{f}} \mu(t)+\frac{t^2}{\alpha^2} \kappa_{\hat{f}} \dot{\mu}(t) \leq \frac{2 t}{\alpha} \kappa_{\hat{f}} \mu_0 t^{-2 \alpha}-\frac{t^2}{\alpha^2} \kappa_{\hat{f}} \mu_0(2 \alpha)\, t^{-2 \alpha-1}=\frac{2 t}{\alpha} \kappa_{\hat{f}} \mu_0 t^{-2 \alpha}-\frac{2 t}{\alpha^2} \kappa_{\hat{f}} \mu_0 t^{-2 \alpha}=0$), and the last inequality holds since $\alpha \geq 2$ and $\mathcal{L}_\beta\left(x^\mu(t), \lambda^*\right)-\mathcal{L}_\beta\left(x^*, \lambda^\mu(t)\right)+2 \kappa_{\hat{f}} \mu(t) \geq 0$.

According to inequality (76), we have $\dot{\hat{\mathcal{V}}}(t) \leq 0, t \geq t_0>0$, i.e., $\mathcal{V}(t)$ is nonincreasing at $t \geq t_0>0$, thus $0<\hat{\mathcal{V}}(t) \leqslant \hat{\mathcal{V}}\left(t_0\right), t \geq t_0>0$.

(I) $\zeta=0, \beta \geq 0$.

(i) If $\alpha \geqslant 2$.

1) From the definition of $\hat{\mathcal{V}}(t)$ in (64) and $\dot{\hat{\mathcal{V}}}(t) \leq 0, t \geq t_0>0$, one has

$$
\begin{aligned}
0 \leq & \frac{1}{2 \mathfrak{l}_{\psi^*}}\left\|\frac{t}{\alpha} \dot{x}^\mu(t)+x^\mu(t)-x^*\right\|^2+\frac{1}{2}\left\|\frac{t}{\alpha} \dot{\lambda}^\mu(t)+\lambda^\mu(t)-\lambda^*\right\|^2 \\
= & \frac{1}{2 \mathfrak{l}_{\psi^*}}\left\|\psi^*\left(u^\mu(t)\right)-x^*\right\|^2+\frac{1}{2}\left\|v^\mu(t)-\lambda^*\right\|^2 \\
\leq & D_{\psi^*}\left(u^\mu(t), u^*\right)+D_h\left(v^\mu(t), \lambda^*\right) \leqslant \hat{\mathcal{V}}\left(t_0\right)<+\infty,
\end{aligned}
$$

which implies that $\sup\limits_{t \in\left[t_0,+\infty\right)}\left\|\frac{t}{\alpha} \dot{\lambda}^\mu(t)+\lambda^\mu(t)\right\|<+\infty$ and $\sup\limits_{t \in\left[t_0,+\infty\right)}\left\|\frac{t}{\alpha} \dot{x}^\mu(t)+x^\mu(t)\right\|<+\infty, \forall x^\mu(t) \in \mathcal{X}$, i.e., $\frac{t}{\alpha} \dot{x}^\mu(t)+x^\mu(t)=\psi^*\left(u^\mu(t)\right)$ and $\frac{t}{\alpha} \dot{\lambda}^\mu(t)+\lambda^\mu(t)=v^\mu(t)$ are bounded.

2) From the definition of $\hat{\mathcal{V}}(t)$ in (64) and $\dot{\hat{\mathcal{V}}}(t) \leq 0, t \geq t_0>0$, we can get

$$
0 \leqslant \frac{t^2}{\alpha^2}\left(\mathcal{L}_\beta\left(x^\mu(t), \lambda^*\right)-\mathcal{L}_\beta\left(x^*, \lambda^\mu(t)\right)+2 \kappa_{\hat{f}} \mu(t)\right) \leqslant \hat{\mathcal{V}}\left(t_0\right)<+\infty,
$$

which means $\mathcal{L}_\beta(x^\mu(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda^\mu(t))+2\kappa_{\hat{f}}\mu(t)\le\frac{\alpha^2\hat{\mathcal{V}}(t_0)}{t^2},\forall t\geqslant t_0>0,x^\mu(t)\in\mathcal{X}$. By using the triangle inequality, one has

$$\begin{aligned}&\left|\mathcal{L}_\beta(x^\mu(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda^\mu(t))\right|\\ \leqslant&\left|\mathcal{L}_\beta(x^\mu(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda^\mu(t))+2\kappa_{\hat{f}}\mu(t)\right|+\left|2\kappa_{\hat{f}}\mu(t)\right|\\ \leqslant&\frac{\alpha^2\hat{\mathcal{V}}(t_0)}{t^2}+\frac{2\kappa_{\hat{f}}\mu_0}{t^{2\alpha}},\forall t\geqslant t_0>0,x^\mu(t)\in\mathcal{X}.\end{aligned}$$

3) From the definition of $\hat{\mathcal{V}}(t)$ in (64) and $\dot{\hat{\mathcal{V}}}(t)\le 0,t\ge t_0>0$, we have

$$0\leqslant\frac{\beta t^2}{2\alpha^2}\left\|Ax^\mu(t)-b\right\|^2\leqslant\hat{\mathcal{V}}(t_0)<+\infty,$$

which implies $\left\|Ax^\mu(t)-b\right\|^2\le\frac{2\alpha^2\hat{\mathcal{V}}(t_0)}{\beta t^2},\forall t\geqslant t_0>0,x^\mu(t)\in\mathcal{X}$. In addition, if $\alpha\geqslant 2,\zeta=0,\beta>0$, it yields to $\dot{\hat{\mathcal{V}}}(t)\leqslant-\frac{\beta t}{2\alpha}\left\|Ax^\mu(t)-b\right\|^2,\forall t\geqslant t_0>0,x^\mu(t)\in\mathcal{X}$, and integrating the inequality above gives

$$\int_{t_0}^{+\infty}t\left\|Ax^\mu(t)-b\right\|^2dt\leqslant\frac{2\alpha}{\beta}\hat{\mathcal{V}}(t_0)<+\infty,\forall t\geqslant t_0>0,x^\mu(t)\in\mathcal{X}.$$

(ii) $\alpha>2$. From (64) with $\alpha>2,\zeta=0,\beta\ge 0$, we can get

$$\dot{\mathcal{V}}(t)\leqslant-t\frac{\alpha-2}{\alpha}\Big(\mathcal{L}_\beta(x^\mu(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda^\mu(t))+2\kappa_{\hat{f}}\mu(t)\Big)\leqslant 0,\forall\, t\geqslant t_0,x^\mu(t)\in\mathcal{X},$$

and integrating the inequality above gives

$$\begin{aligned}&\int_{t_0}^{+\infty}t\Big(\mathcal{L}_\beta(x^\mu(t),\lambda^*)-\mathcal{L}_\beta(x^*,\lambda^\mu(t))+2\kappa_{\hat{f}}\mu(t)\Big)dt\\ \leqslant&\frac{\alpha}{\alpha-2}\hat{\mathcal{V}}(t_0)<+\infty,\forall t\geqslant t_0>0,x^\mu(t)\in\mathcal{X}.\end{aligned}$$

(II) $\zeta>0,\beta\geqslant 0$.
(i) $\alpha\geqslant 2$.
1) According to the definition of $\hat{\mathcal{V}}(t)$ in (64) and $\dot{\hat{\mathcal{V}}}(t)\le 0,t\ge t_0>0$, one has

$$\begin{aligned}&\sup_{t\in[t_0,+\infty)}\left\|\frac{t}{\alpha}\dot{x}^\mu(t)+x^\mu(t)-x^*\right\|\leqslant\sqrt{2\hat{\mathcal{V}}(t_0)}<+\infty,\forall x^\mu(t)\in\mathcal{X}\\ &\sup_{t\in[t_0,+\infty)}\left\|\frac{t}{\alpha}\dot{\lambda}^\mu(t)+\lambda^\mu(t)-\lambda^*\right\|\leqslant\sqrt{2\hat{\mathcal{V}}(t_0)}<+\infty,\\ &\sup_{t\in[t_0,+\infty)}\left\|x^\mu(t)-x^*\right\|\leqslant\sqrt{\frac{2}{\zeta}\hat{\mathcal{V}}(t_0)}<+\infty,\forall x^\mu(t)\in\mathcal{X}\\ &\sup_{t\in[t_0,+\infty)}\left\|\lambda^\mu(t)-\lambda^*\right\|\leqslant\sqrt{\frac{2}{\zeta}\hat{\mathcal{V}}(t_0)}<+\infty,\end{aligned}$$

which means that $x^{\mu}(t), \forall\, t \geqslant t_0, x^{\mu}(t) \in \mathcal{X}$ and $\lambda^{\mu}(t)$ are bounded. In addition, by using the triangle inequality, we obtain

$$\begin{aligned}
t\left\|\dot{x}^{\mu}(t)\right\| \leqslant& \alpha\left\|\frac{t}{\alpha}\dot{x}^{\mu}(t)+x^{\mu}(t)-x^{*}\right\|+\alpha\left\|x^{\mu}(t)-x^{*}\right\| \\
\leqslant& a\left(\sqrt{\zeta}+1\right)\sqrt{\frac{2}{\zeta}\hat{\mathcal{V}}(t_0)}<+\infty, \forall\, t \geqslant t_0>0, x^{\mu}(t)\in\mathcal{X}; \\
t\left\|\dot{\lambda}^{\mu}(t)\right\| \leqslant& \alpha\left\|\frac{t}{\alpha}\dot{\lambda}^{\mu}(t)+\lambda^{\mu}(t)-\lambda^{*}\right\|+\alpha\left\|\lambda^{\mu}(t)-\lambda^{*}\right\| \\
\leqslant& a\left(\sqrt{\zeta}+1\right)\sqrt{\frac{2}{\zeta}\hat{\mathcal{V}}(t_0)}<+\infty, \forall\, t \geqslant t_0,
\end{aligned}$$

it means $\left\|\dot{x}^{\mu}(t)\right\| = O\left(\frac{1}{t}\right), \forall t \geqslant t_0, x^{\mu}(t) \in \mathcal{X}$ and $\left\|\dot{\lambda}^{\mu}(t)\right\| = O\left(\frac{1}{t}\right), \forall t \geqslant t_0$.

2) From (64) with $\alpha \geq 2, \zeta > 0, \beta \geq 0$, we can get $\dot{\mathrm{V}}(t) \leqslant -\frac{\zeta t}{\alpha}\left\|\dot{x}^{\mu}(t)\right\|^2 - \frac{\zeta t}{\alpha}\left\|\dot{\lambda}^{\mu}(t)\right\|^2, \forall\, t \geqslant t_0 > 0, x^{\mu}(t) \in \mathcal{X}$, and integrating the inequality above, we can obtain for any $t \geq t_0 > 0$

$$\begin{aligned}
\int_{t_0}^{+\infty} t\left\|\dot{x}^{\mu}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\hat{\mathcal{V}}(t_0) < +\infty, \\
\int_{t_0}^{+\infty} t\left\|\dot{\lambda}^{\mu}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\hat{\mathcal{V}}(t_0) < +\infty.
\end{aligned}$$

3) In addition, together with (64) and $\dot{\hat{\mathcal{V}}}(t) \leq 0, t \geq t_0 > 0$, we can get

$$0 \leqslant \frac{t^2}{\alpha^2}\left(\mathcal{L}_{\beta}\left(x^{\mu}(t), \lambda^{*}\right) - \mathcal{L}_{\beta}\left(x^{*}, \lambda^{\mu}(t)\right) + 2\kappa_{\hat{f}}\mu(t)\right) \leqslant \hat{\mathcal{V}}(t_0) < +\infty,$$

which means $\mathcal{L}_{\beta}\left(x^{\mu}(t), \lambda^{*}\right) - \mathcal{L}_{\beta}\left(x^{*}, \lambda^{\mu}(t)\right) + 2\kappa_{\hat{f}}\mu(t) \leq \frac{\alpha^2\hat{\mathcal{V}}(t_0)}{t^2}, \forall\, t \geqslant t_0 > 0, x^{\mu}(t) \in \mathcal{X}$. By using the triangle inequality, one has

$$\begin{aligned}
&\left|\mathcal{L}_{\beta}\left(x^{\mu}(t), \lambda^{*}\right) - \mathcal{L}_{\beta}\left(x^{*}, \lambda^{\mu}(t)\right)\right| \\
\leqslant& \left|\mathcal{L}_{\beta}\left(x^{\mu}(t), \lambda^{*}\right) - \mathcal{L}_{\beta}\left(x^{*}, \lambda^{\mu}(t)\right) + 2\kappa_{\hat{f}}\mu(t)\right| + \left|2\kappa_{\hat{f}}\mu(t)\right| \\
\leqslant& \frac{\alpha^2\hat{\mathcal{V}}(t_0)}{t^2} + \frac{2\kappa_{\hat{f}}\mu_0}{t^{2\alpha}}, \forall\, t \geqslant t_0 > 0, x^{\mu}(t) \in \mathcal{X}.
\end{aligned}$$

From the first, third and fourth equalities in SAPDMD (60), for every $t \geq t_0$, we have

$$\begin{aligned}
\lambda^{\mu}(t) - \lambda^{\mu}(t_0) =& \int_{t_0}^{t} \dot{\lambda}^{\mu}(s)\, ds \\
=& \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{s}{\alpha}\left(A\nabla\psi^{*}\left(u^{\mu}(s)\right) - b\right) ds \;-\frac{1}{\zeta}\int_{t_0}^{t} \dot{v}^{\mu}(s)\, ds \\
=& \frac{1}{\zeta}\int_{t_0}^{t} \tfrac{s}{\alpha}\left(A\nabla\psi^{*}\left(u^{\mu}(s)\right) - b\right) ds \\
&+ \frac{t_0}{\zeta\alpha}\dot{\lambda}^{\mu}(t_0) + \frac{1}{\zeta}\lambda^{\mu}(t_0) - \frac{t}{\zeta\alpha}\dot{\lambda}^{\mu}(t) - \frac{1}{\zeta}\lambda^{\mu}(t)
\end{aligned}$$

$$
\begin{aligned}
&\Rightarrow \frac{t}{\zeta\alpha}\dot{\lambda}^{\mu}(t)+\frac{1+\zeta}{\zeta}\lambda^{\mu}(t)-\frac{t_0}{\zeta\alpha}\dot{\lambda}^{\mu}(t_0)-\frac{1+\zeta}{\zeta}\lambda^{\mu}(t_0)\\
&=\frac{1}{\zeta}\int_{t_0}^{t}\ \tfrac{s}{\alpha}\left(Ax^{\mu}(s)-b\right)ds\ \ +\frac{1}{\zeta}\int_{t_0}^{t}\ \tfrac{s^2}{\alpha^2}d\left(Ax^{\mu}(s)-b\right)ds\\
&=\frac{1}{\zeta}\int_{t_0}^{t}\ \tfrac{s}{\alpha}\left(Ax^{\mu}(s)-b\right)ds\ \ +\frac{t^2}{\zeta\alpha^2}\left(Ax^{\mu}(t)-b\right)\\
&\quad-\frac{t_0^2}{\zeta\alpha^2}\left(Ax^{\mu}(t_0)-b\right)-\frac{1}{\zeta}\int_{t_0}^{t}\ \tfrac{2s}{\alpha^2}\left(Ax^{\mu}(s)-b\right)ds\\
&=\frac{t^2}{\zeta\alpha^2}\left(Ax^{\mu}(t)-b\right)-\frac{t_0^2}{\zeta\alpha^2}\left(Ax^{\mu}(t_0)-b\right)\\
&\quad+\int_{t_0}^{t}\frac{s^2}{\zeta\alpha^2}\left(\frac{\alpha-2}{s}\right)\ \left(Ax^{\mu}(s)-b\right)ds,
\end{aligned}
\tag{77}
$$

where the second equality holds from the fourth equality in SAPDMD (60) $\dot{\lambda}^{\mu}(t)=-\frac{1}{\zeta}\dot{v}^{\mu}(t)+v^{\mu}$, the third equality is satisfied due to $v^{\mu}(t)=\frac{t}{\alpha}\dot{\lambda}^{\mu}(t)+\lambda^{\mu}(t)$, i.e., the third equality in SAPDMD (60), and the fourth equality holds because of $\nabla\psi^{*}(u^{\mu}(t))=x^{\mu}(t)+\frac{t}{\alpha}\dot{x}^{\mu}(t)$, i.e., the first equality in SAPDMD (60).

It follows from (77) that, for any $t\geqslant t_0>0$,

$$
\left\|\frac{t^2}{\zeta\alpha^2}\left(Ax^{\mu}(t)-b\right)+\int_{t_0}^{t}\left(\frac{\alpha-2}{s}\right)\ \tfrac{s^2}{\zeta\alpha^2}\left(Ax^{\mu}(s)-b\right)ds\ \right\|\leqslant\mathcal{C}_1,
\tag{78}
$$

where $\mathcal{C}_1=\sup\limits_{t\in[t_0,+\infty)}\left\{\frac{t}{\zeta\alpha}\left\|\dot{\lambda}^{\mu}(t)\right\|+\frac{t_0}{\zeta\alpha}\left\|\dot{\lambda}^{\mu}(t_0)\right\|+\frac{1+\zeta}{\zeta}\left\|\lambda^{\mu}(t)-\lambda^{\mu}(t_0)\right\|+\frac{t_0^2}{\alpha^2}\left\|Ax^{\mu}(t_0)-b\right\|\right\}<+\infty$. Setting

$$
g(t)=\frac{t^2}{\zeta\alpha^2}\left(Ax^{\mu}(t)-b\right),a(t)=\frac{\alpha-2}{\alpha^2},\forall t\geqslant t_0>0,
$$

and applying **Lemma 12** to get that

$$
\frac{t^2}{\zeta\alpha^2}\left\|Ax^{\mu}(t)-b\right\|\leqslant 2\mathcal{C}_1\Rightarrow\left\|Ax^{\mu}(t)-b\right\|\leqslant\frac{2\zeta\alpha^2\mathcal{C}_1}{t^2},\forall t\geqslant t_0>0,x^{\mu}(t)\in\mathcal{X}.
\tag{79}
$$

Since $\mathcal{L}_\beta(x^{\mu}(t),\lambda^{*})-\mathcal{L}_\beta(x^{*},\lambda^{\mu}(t))=f(x^{\mu}(t))-f(x^{*})+\frac{\beta}{2}\left\|Ax^{\mu}(t)-b\right\|^2+\left(Ax^{\mu}(t)-b\right)^{T}\lambda^{*}$, by using (79), we have

$$
\begin{aligned}
&\left|f(x^{\mu}(t))-f(x^{*})\right|\\
&\leqslant\left|\mathcal{L}_\beta(x^{\mu}(t),\lambda^{*})-\mathcal{L}_\beta(x^{*},\lambda^{\mu}(t))\right|+\|\lambda^{*}\|\left\|Ax^{\mu}(t)-b\right\|+\frac{\beta}{2}\left\|Ax^{\mu}(t)-b\right\|^2\\
&\leqslant\frac{\alpha^2}{t^2}\left(\hat{\mathcal{V}}(t_0)+2\zeta\|\lambda^{*}\|\mathcal{C}_1\right)+\ \frac{2\beta\zeta^2\alpha^4\mathcal{C}_1^2}{t^4}\ ,\forall t\geqslant t_0>0,x^{\mu}(t)\in\mathcal{X}.
\end{aligned}
$$

(ii) $\alpha>2$. From (64) with $\alpha>2,\zeta=0,\beta\geq 0$, we can get

$$
\dot{\mathrm{V}}(t)\leqslant -t\frac{\alpha-2}{\alpha}\left(\mathcal{L}_\beta(x^{\mu}(t),\lambda^{*})-\mathcal{L}_\beta(x^{*},\lambda^{\mu}(t))+2\kappa_{\hat{f}}\mu(t)\right)\leqslant 0,\forall\ t\geqslant t_0,x^{\mu}(t)\in\mathcal{X},
$$

and integrating the inequality above gives

$$\begin{aligned}&\int_{t_0}^{+\infty} t\Big(\mathcal{L}_\beta\left(x^\mu\left(t\right),\lambda^*\right)-\mathcal{L}_\beta\left(x^*,\lambda^\mu\left(t\right)\right)+2\kappa_{\hat{f}}\mu\left(t\right)\Big)dt\\ \leqslant&\frac{\alpha}{\alpha-2}\hat{\mathcal{V}}\left(t_0\right)<+\infty,\forall\, t\geqslant t_0>0,x^\mu\left(t\right)\in\mathcal{X}.\end{aligned}$$

Thus, the proof is completed. ■

Similar to the smooth case, the SAPDMD (60) can also be used to address the DCCP (12) and DEMO (13) with nonsmooth convex objections.

4.4 SAPDMD for DCCP in the nonsmooth case

Based on the SAPDMD (60), a smoothing accelerated dirtibuted primal-dual mirror dynamical (SADPDMD) approach for DCCP (12) with nonsmooth convex objective functions is given by:

$$\begin{cases}\dot{x}^\mu\left(t\right)=\frac{\alpha}{t}\left(\nabla\psi^*\left(u^\mu\left(t\right)\right)-x^\mu\left(t\right)\right),\\ \dot{u}^\mu\left(t\right)=-\frac{t}{\alpha}\Big(\nabla_x\hat{f}\left(x^\mu\left(t\right),\mu\left(t\right)\right)+\beta Lx^\mu\left(t\right)+L\left(v^\mu\left(t\right)\right)\Big)-\zeta\dot{x}^\mu\left(t\right),\\ \dot{\lambda}^\mu\left(t\right)=\frac{\alpha}{t}\left(v^\mu\left(t\right)-\lambda^\mu\left(t\right)\right),\dot{v}^\mu\left(t\right)=\frac{t}{\alpha}L\nabla\psi^*\left(u^\mu\left(t\right)\right)-\zeta\dot{\lambda}^\mu\left(t\right),\\ x^\mu\left(t_0\right)=x_0^\mu,\nabla\psi^*\left(u^\mu\left(t_0\right)\right)=x_0^\mu,\lambda^\mu\left(t_0\right)=\lambda_0^\mu,v^\mu\left(t_0\right)=v_0^\mu,\mu_0=\bar{\mu}>0,\end{cases}\tag{80}$$

where $t\geq t_0>0$, $\beta\geq 0$, $\alpha\geq 2$, and $\mu\left(t\right)\leq\frac{\mu_0}{t^{2\alpha}}$.

4.5 The accelerated convergence of the SADPDMD

Now, let's discuss accelerated convergence properties of SADPDMD (80) with the help of Lyapunov analysis tool.

Theorem 18. *Suppose that **Assumption 3.2** holds and the objective function is nonsmooth. Let $\left(x^\mu\left(t\right),\lambda^\mu\left(t\right)\right)$ and $\left(x^*,\lambda^*\right)$ be a solution trajectory and an optimal solution of SADPDMD (80), respectively. Let $\hat{\mathsf{L}}_\beta\left(x^\mu\left(t\right),\lambda^\mu\left(t\right)\right)=\hat{f}\left(x^\mu\left(t\right),\mu\left(t\right)\right)+\frac{\beta}{2}\left(x^\mu\left(t\right)\right)^TL\left(x^\mu\left(t\right)\right)+\left(\lambda^\mu\left(t\right)\right)^TLx^\mu\left(t\right)$, $\mathsf{L}_\beta\left(x^\mu\left(t\right),\lambda^\mu\left(t\right)\right)=f\left(x^\mu\left(t\right)\right)+\frac{\beta}{2}\left(x^\mu\left(t\right)\right)^TL\left(x^\mu\left(t\right)\right)+\left(\lambda^\mu\left(t\right)\right)^TLx^\mu\left(t\right)$, and constructing a candidate smoothing Lyapunov function as*

$$\hat{\mathrm{V}}=\hat{\mathrm{V}}_1\left(t\right)+\hat{\mathrm{V}}_2\left(t\right)+\hat{\mathrm{V}}3\left(t\right),\tag{81}$$

with

$$\begin{cases}\hat{\mathrm{V}}_1\left(t\right)=\frac{t^2}{\alpha^2}\Big(\hat{\mathsf{L}}_\beta\left(x^\mu\left(t\right),\lambda^*\right)-\hat{\mathsf{L}}_\beta\left(x^*,\lambda^\mu\left(t\right)\right)+4\kappa_{\hat{f}}\mu\left(t\right)\Big)\\ \qquad\quad=\frac{t^2}{\alpha^2}\Big(\hat{f}\left(x^\mu\left(t\right),\mu\left(t\right)\right)+\frac{\beta}{2}\left(x^\mu\left(t\right)\right)^TL\left(x^\mu\left(t\right)\right)\\ \qquad\quad+\left(\lambda^*\right)^TLx^\mu\left(t\right)+\hat{f}\left(x^*,\mu\left(t\right)\right)+4\kappa_{\hat{f}}\mu\left(t\right)\Big)\\ \hat{\mathrm{V}}_2\left(t\right)=D_{\psi^*}\left(u^\mu\left(t\right),u^*\right)+\frac{\zeta}{2}\left\|u^\mu\left(t\right)-u^*\right\|^2\\ \qquad\quad=\psi^*\left(u^\mu\left(t\right)\right)-\psi^*\left(u^*\right)-\nabla\psi^*\left(u^*\right)^T\left(u^\mu\left(t\right)-u^*\right),\\ \hat{\mathrm{V}}_3\left(t\right)=D_h\left(v^\mu\left(t\right),v^*\right)+\frac{\zeta}{2}\left\|\lambda^\mu\left(t\right)-\lambda^*\right\|^2\\ \qquad\quad=\frac{1}{2}\left\|v^\mu\left(t\right)-v^*\right\|^2+\frac{\zeta}{2}\left\|\lambda^\mu\left(t\right)-\lambda^*\right\|^2.\end{cases}$$

where $u^* = \nabla\psi(x^*)$, $v^* = \lambda^*$. *The following statements are true:*

(I) $\zeta = 0, \beta \geq 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, *we have*

1) $\frac{t}{\alpha}\dot{x}^\mu(t) + x^\mu(t) = \psi^*(u^\mu(t))$ *and* $\frac{t}{\alpha}\dot{\lambda}^\mu(t) + \lambda^\mu(t) = v^\mu(t)$ *of SADPDMD* (80) *are bounded;*

2) One has

$$\left|\mathsf{L}_\beta(x^\mu(t), \lambda^*) - \mathsf{L}_\beta(x^*, \lambda^\mu(t))\right| \leqslant \frac{\alpha^2\hat{\mathrm{V}}(t_0)}{t^2} + \frac{2\kappa_{\hat{f}}\mu_0}{t^{2\alpha}}; \tag{82}$$

3) if, in addition, $\beta > 0$,

$$x^\mu(t)^T Lx^\mu(t) \leq \frac{2\alpha^2\hat{\mathrm{V}}(t_0)}{\beta t^2}; \tag{83a}$$

$$\int_{t_0}^{+\infty} tx^\mu(t)^T Lx^\mu(t)dt < +\infty; \tag{83b}$$

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we have*

$$\int_{t_0}^{+\infty} t\Big(\mathsf{L}_\beta(x^\mu(t), \lambda^*) - \mathsf{L}_\beta(x^*, \lambda^\mu(t)) + 2\kappa_{\hat{f}}\mu(t)\Big)dt < +\infty; \tag{84}$$

(II) $\zeta > 0, \beta \geqslant 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, *one has*

1) $x^\mu(t)$ *and* $\lambda^\mu(t)$ *are bounded;*

$$\|\dot{x}^\mu(t)\| = O\left(\frac{1}{t}\right), \left\|\dot{\lambda}^\mu(t)\right\| = O\left(\frac{1}{t}\right); \tag{85}$$

2)

$$\int_{t_0}^{+\infty} t\left\|\dot{x^\mu}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\hat{\mathrm{V}}(t_0) < +\infty; \tag{86a}$$

$$\int_{t_0}^{+\infty} t\left\|\dot{\lambda^\mu}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\hat{\mathrm{V}}(t_0) < +\infty; \tag{86b}$$

3)

$$\left|\mathsf{L}_\beta(x^\mu(t), \lambda^*) - \mathsf{L}_\beta(x^*, \lambda^\mu(t))\right| \leqslant \frac{\alpha^2\hat{\mathrm{V}}(t_0)}{t^2} + \frac{2\kappa_{\hat{f}}\mu_0}{t^{2\alpha}}; \tag{87a}$$

$$\sqrt{x^\mu(t)^T Lx^\mu(t)} \leqslant \frac{2\zeta\alpha^2\mathcal{C}_2}{t^2};; \tag{87b}$$

$$|f(x^\mu(t)) - f(x^*)| \leqslant \frac{\alpha^2}{t^2}\Big(\hat{\mathrm{V}}(t_0) + 2\zeta\|\lambda^*\|\mathcal{C}_2\Big) + \frac{2\beta\zeta^2\alpha^4\mathcal{C}_2^2}{t^4}; \tag{87c}$$

where $\mathcal{C}_2 = \sup\limits_{t\in[t_0,+\infty)}\left\{\frac{t}{\zeta\alpha}\left\|\dot{\lambda}^\mu(t)\right\| + \frac{t_0}{\zeta\alpha}\left\|\dot{\lambda}^\mu(t_0)\right\| + \frac{1+\zeta}{\zeta}\|\lambda^\mu(t) - \lambda^\mu(t_0)\| + \frac{t_0^2}{\alpha^2}\|Lx^\mu(t_0)\|\right\} < +\infty$.

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x(t) \in \mathcal{X}$, *we have*

$$\int_{t_0}^{+\infty} t\left(\mathsf{L}_\beta\left(x^\mu(t), \lambda^*\right) - \mathsf{L}_\beta\left(x^*, \lambda^\mu(t)\right) + 2\kappa_{\hat{f}}\mu(t)\right)dt < +\infty. \tag{88}$$

Proof Following similar steps as the proof in **Theorem 17**, 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 (13) where the objective function is nonsmooth, a smoothing accelerated distributed mirror dynamical (SADMD) approach inspired by the SAPDMD (60) and distributed consensus theorem is given by

$$\begin{cases} \dot{x}^\mu(t) = \frac{\alpha}{t}\left(\nabla\psi^*\left(u^\mu(t)\right) - x^\mu(t)\right), \\ \dot{u}^\mu(t) = -\frac{t}{\alpha}\left(\nabla_x \hat{f}\left(x^\mu(t), \mu(t)\right) + \bar{A}^T v^\mu(t)\right) - \zeta\dot{x}^\mu(t), \\ \dot{\lambda}^\mu(t) = \frac{\alpha}{t}\left(v^\mu(t) - \lambda^\mu(t)\right), \\ \dot{v}^\mu(t) = \frac{t}{\alpha}\left(\bar{A}\psi^*\left(u^\mu(t)\right) - d - \beta L\lambda^\mu(t) + Lz^\mu(t)\right) - \zeta\dot{\lambda}^\mu(t), \\ \dot{y}^\mu(t) = \frac{\alpha}{t}\left(z^\mu(t) - y^\mu(t)\right), \dot{z}(t) = -\frac{t}{\alpha}Lv^\mu(t) - \zeta\dot{y}^\mu(t), \\ x^\mu(t_0) = x_0^\mu, u^\mu(t_0) = u_0^\mu \text{ with } \nabla\psi^*(u_0) = x_0^\mu \in \mathcal{X}, \\ \lambda^\mu(t_0) = \lambda_0^\mu, v^\mu(t_0) = v_0^\mu, y^\mu(t_0) = y_0^\mu, z^\mu(t_0) = z_0^\mu, \mu_0 = \bar{\mu} > 0, \end{cases} \tag{89}$$

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

The accelerated convergence properties of SADMD (89) will be demonstrated in the following theorem.

Theorem 19. *Suppose that* ***Assumption 3.2*** *holds, except that the objective function is nonsmooth. Let* $\left(x^\mu(t), \lambda^\mu(t), y^\mu(t)\right)$ *and* $\left(x^*, \lambda^*, y^*\right)$ *be a solution trajectory and an optimal solution of SADMD* (89)*, respectively. Then, for any initial values* $\left(x^\mu(t_0), u^\mu(t_0), \lambda^\mu(t_0),\right.$ $\left.v^\mu(t_0), y^\mu(t_0), z^\mu(t_0)\right) \in \mathcal{X} \times \mathbb{R}^{\sum_{i=1}^n p_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ *and let* $\hat{\mathbb{L}}_\beta\left(x^\mu(t), \lambda^\mu(t), y^\mu(t)\right)$ *be* $\hat{f}\left(x^\mu(t), \mu(t)\right) - \frac{\beta}{2}\lambda^\mu(t)^T L\lambda^\mu(t) + \lambda^\mu(t)^T\left(\bar{A}x^\mu(t) - d + Ly^\mu(t)\right)$, $\mathbb{L}_\beta\left(x^\mu(t), \lambda^\mu(t), y^\mu(t)\right) =$ $f\left(x^\mu(t)\right) - \frac{\beta}{2}\lambda^\mu(t)^T L\lambda^\mu(t) + \lambda^\mu(t)^T\left(\bar{A}x^\mu(t) - d + Ly^\mu(t)\right)$ *and constructing a smoothing Lyapunov function as follows:*

$$\hat{\mathbb{V}}(t) = \hat{\mathbb{V}}_1(t) + \hat{\mathbb{V}}_2(t) + \hat{\mathbb{V}}_3(t) + \hat{\mathbb{V}}_4(t), \tag{90}$$

with

$$\begin{cases} \hat{\mathbb{V}}_\beta(t) = \frac{t^2}{\alpha^2}\left(\hat{\mathbb{L}}_\beta\left(x^\mu(t)(t), y^*, \lambda^*\right) - \hat{\mathbb{L}}_\beta\left(x^*, y^*, \lambda^\mu(t)\right) + 4\kappa_{\hat{f}}\mu(t)\right), \\ \qquad = \frac{t^2}{\alpha^2}\left(\hat{f}\left(x^\mu(t), \mu(t)\right) - \hat{f}\left(x^*, \mu(t)\right) + \frac{\beta}{2}\lambda^\mu(t)^T L\lambda^\mu(t),\right. \\ \qquad \left. + 4\kappa_{\hat{f}}\mu(t) + (\lambda^*)^T\left(\bar{A}x^\mu(t) - d - Ly^*\right)\right), \\ \hat{\mathbb{V}}_2(t) = D_{\psi^*}\left(u^\mu(t), u^*\right) + \frac{\zeta}{2}\left\|x^\mu(t) - x^*\right\|^2 \\ \qquad = \psi^*\left(u^\mu(t)\right) - \psi^*\left(u^*\right) - \nabla\psi^*\left(u^*\right)^T\left(u^\mu(t) - u^*\right) + \frac{\zeta}{2}\left\|x^\mu(t) - x^*\right\|^2, \\ \hat{\mathbb{V}}_3(t) = D_h\left(v^\mu(t)(t), v^*\right) + \frac{\zeta}{2}\left\|\lambda^\mu(t) - \lambda^*\right\|^2 = \frac{1}{2}\left\|v^\mu(t) - \lambda^*\right\|^2 + \frac{\zeta}{2}\left\|x^\mu(t) - x^*\right\|^2, \\ \hat{\mathbb{V}}_4(t) = D_h\left(z^\mu(t), z^*\right) + \frac{\zeta}{2}\left\|y^\mu(t) - y^*\right\|^2 = \frac{1}{2}\left\|z^\mu(t) - y^*\right\|^2 + \frac{\zeta}{2}\left\|y^\mu(t) - y^*\right\|^2, \end{cases}$$

where $u^* = \nabla\psi(x^*)$, $v^* = \lambda^*$ *and* $z^* = y^*$. *The following statements are true.*

(I) $\zeta = 0$, $\beta \geq 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, *we have*

1) $\frac{t}{\alpha}\dot{x^\mu}(t) + x^\mu(t) = \psi^*(u^\mu(t))$, $\frac{t}{\alpha}\dot{\lambda^\mu}(t) + \lambda^\mu(t) = v(t)$ *and* $\frac{t}{\alpha}\dot{y^\mu}(t) + y^\mu(t) = z^\mu(t)$ *of SADMD* (89) *are bounded;*

2) One has

$$|\mathbb{L}_\beta(x^\mu(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda^\mu(t))| \leqslant \frac{\alpha^2\hat{\mathbb{V}}(t_0)}{t^2} + \frac{2\kappa_{\hat{f}}\mu_0}{t^{2\alpha}}; \tag{91}$$

3) if, in addition, $\beta \geq 0$, *one has*

$$\lambda^\mu(t)^T L\lambda^\mu(t) \leq \frac{2\alpha^2\hat{\mathbb{V}}(t_0)}{\beta t^2}; \tag{92a}$$

$$\int_{t_0}^{+\infty} t\lambda^\mu(t)^T L\lambda^\mu(t)dt < +\infty; \tag{92b}$$

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, *we can get*

$$\int_{t_0}^{+\infty} t\Big(\mathbb{L}_\beta(x^\mu(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda^\mu(t)) + 2\kappa_{\hat{f}}\mu(t)\Big)dt < +\infty; \tag{93}$$

(II) $\zeta > 0$ *and* $\beta \geq 0$.

(i) $\alpha \geqslant 2$. *For any* $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, *one has*

1) $x^\mu(t)$, $y^\mu(t)$ *and* $\lambda^\mu(t)$ *of SADMD* (89) *are bounded;*

$$\|\dot{x}^\mu(t)\| = O\left(\frac{1}{t}\right), \|\dot{y}^\mu(t)\| = O\left(\frac{1}{t}\right), \left\|\dot{\lambda}^\mu(t)\right\| = O\left(\frac{1}{t}\right); \tag{94}$$

2)

$$\int_{t_0}^{+\infty} t\left\|\dot{x^\mu}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\hat{\mathbb{V}}(t_0) < +\infty; \tag{95a}$$

$$\int_{t_0}^{+\infty} t\|\dot{y}(t)\|^2 dt \leqslant \frac{\alpha}{\zeta}\hat{\mathbb{V}}(t_0) < +\infty; \tag{95b}$$

$$\int_{t_0}^{+\infty} t\left\|\dot{\lambda}(t)\right\|^2 dt \leqslant \frac{\alpha}{\zeta}\hat{\mathbb{V}}(t_0) < +\infty; \tag{95c}$$

3)

$$|\mathbb{L}_\beta(x^\mu(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda^\mu(t))| \leqslant \frac{\alpha^2\hat{\mathbb{V}}(t_0)}{t^2} + \frac{2\kappa_{\hat{f}}\mu_0}{t^{2\alpha}}; \tag{96}$$

(ii) $\alpha > 2$. *For any* $t \geqslant t_0 > 0, x^\mu(t) \in \mathcal{X}$, *we have*

$$\int_{t_0}^{+\infty} t\Big(\mathbb{L}_\beta(x^\mu(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda^\mu(t)) + 2\kappa_{\hat{f}}\mu(t)\Big)dt < +\infty. \tag{97}$$

Proof The time derivative of $\hat{\mathbb{V}}(t)$ with $\bar{A}x^* = d + Ly^*$, $v^* = \lambda^*$, $y^* = z^*$, $\nabla\psi^*(u^*) = x^*$ and $L\lambda^* = 0$ is

$$
\begin{aligned}
\dot{\hat{\mathbb{V}}}(t) =& \dot{\hat{\mathbb{V}}}_1(t) + \dot{\hat{\mathbb{V}}}_2(t) + \dot{\hat{\mathbb{V}}}_3(t) + \dot{\hat{\mathbb{V}}}_4(t) \\
\leq& \frac{2t}{\alpha^2}\Big(\hat{f}(x^\mu(t), \mu(t)) - \hat{f}(x^*, \mu(t)) + (\lambda^*)^T \bar{A}(x^\mu(t) - x^*) + 4\kappa_{\hat{f}}\mu(t) \\
&+ \frac{\beta}{2}\lambda^\mu(t)^T L\lambda^\mu(t)\Big) - \frac{t\beta}{\alpha}(L\lambda^\mu(t))^T\lambda^\mu(t) + \frac{4t}{\alpha}\kappa_{\hat{f}}\mu(t) + \frac{2t^2}{\alpha^2}\kappa_{\hat{f}}\dot{\mu}(t) \\
&- \frac{t}{\alpha}\Big(\hat{f}(x^\mu(t), \mu(t)) - \hat{f}(x^*, \mu(t)) + (\lambda^*)^T \bar{A}(x^\mu(t) - x^*) + 4\kappa_{\hat{f}}\mu(t)\Big) \\
&- \frac{t\zeta}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2 - \frac{t\zeta}{\alpha}\|\dot{y}^\mu(t)\|^2 - \frac{t\zeta}{\alpha}\|\dot{x}^\mu(t)\|^2 \\
\leq& -\frac{(\alpha-2)t}{\alpha^2}\Big(\hat{f}(x^\mu(t), \mu(t)) - \hat{f}(x^*, \mu(t)) + (\lambda^*)^T \bar{A}(x^\mu(t) - x^*) \\
&+ 4\kappa_{\hat{f}}\mu(t) + \frac{\beta}{2}\lambda^\mu(t)^T L\lambda^\mu(t)\Big) - \frac{t\beta}{2\alpha}\lambda^\mu(t)^T L\lambda^\mu(t) \\
&- \frac{t\zeta}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2 - \frac{t\zeta}{\alpha}\|\dot{y}^\mu(t)\|^2 - \frac{t\zeta}{\alpha}\|\dot{x}^\mu(t)\|^2 \\
=& -\frac{(\alpha-2)t}{\alpha^2}\Big(\hat{\mathbb{L}}_\beta(x^\mu(t), y^*, \lambda^*) - \hat{\mathbb{L}}_\beta(x^*, y^*, \lambda^\mu(t)) + 4\kappa_{\hat{f}}\mu(t)\Big) \\
&- \frac{t\zeta}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2 - \frac{t\zeta}{\alpha}\|\dot{y}^\mu(t)\|^2 - \frac{t\zeta}{\alpha}\|\dot{x}^\mu(t)\|^2 - \frac{\beta t}{2\alpha}\lambda^\mu(t)^T L\lambda^\mu(t) \\
\leqslant& -\frac{(\alpha-2)t}{\alpha^2}\Big(\mathbb{L}_\beta(x^\mu(t), y^*, \lambda^*) - \mathbb{L}_\beta(x^*, y^*, \lambda^\mu(t)) + 2\kappa_{\hat{f}}\mu(t)\Big) \\
&- \frac{t\zeta}{\alpha}\left\|\dot{\lambda}^\mu(t)\right\|^2 - \frac{t\zeta}{\alpha}\|\dot{y}^\mu(t)\|^2 - \frac{t\zeta}{\alpha}\|\dot{x}^\mu(t)\|^2 - \frac{\beta t}{2\alpha}\lambda^\mu(t)^T L\lambda^\mu(t) \\
\leq& 0,
\end{aligned}
\tag{98}
$$

where the first inequality holds because of $\nabla_\mu\hat{f}(x^\mu(t), \mu(t)) + \nabla_\mu\hat{f}(x^*, \mu(t)) + 2\kappa_{\hat{f}} \geq 0$, $\dot{\mu}(t) \leq 0$, the second inequality is satisfied due to the convexity of $\hat{f}(x^\mu(t), \mu(t))$ of $x^\mu(t) \in \mathcal{X}$ with any fixed $\mu(t) \in [0, \bar{\mu}]$ and the property of $\mu(t) \leq \frac{\mu_0}{t^{2\alpha}}$ (i.e., $\dot{\mu}(t) \leq -\mu_0(2\alpha)t^{(-2\alpha-1)}$ and $\frac{t^2}{\alpha^2}\kappa_{\hat{f}} \geq 0$ implies $\frac{2t}{\alpha}\kappa_{\hat{f}}\mu(t) + \frac{t^2}{\alpha^2}\kappa_{\hat{f}}\dot{\mu}(t) \leq \frac{2t}{\alpha}\kappa_{\hat{f}}\mu_0 t^{-2\alpha} - \frac{t^2}{\alpha^2}\kappa_{\hat{f}}\mu_0(2\alpha)t^{-2\alpha-1} = \frac{2t}{\alpha}\kappa_{\hat{f}}\mu_0 t^{-2\alpha} - \frac{2t}{\alpha^2}\kappa_{\hat{f}}\mu_0 t^{-2\alpha} = 0$), the third inequality holds due to $\hat{f}(x^\mu(t), \mu(t)) + \kappa_{\hat{f}}\mu(t) \geqslant f(x^\mu(t))$, i.e., $\hat{\mathbb{L}}_\beta(x^\mu(t), y^*, \lambda^*) + \kappa_{\hat{f}} \geqslant \mathbb{L}_\beta(x^\mu(t), y^*, \lambda^*)$ and $(x^*, \mu(t)) + \kappa_{\hat{f}}\mu(t) \geqslant f(x^*)$ in Definition 6 (iv), i.e., $\hat{\mathbb{L}}_\beta(x^\mu(t), y^*, \lambda^*) + \kappa_{\hat{f}}\mu(t) \geqslant \mathbb{L}_\beta(x^\mu(t), y^*, \lambda^*)$ and $\mathbb{L}_\beta(x^*, y^*, \lambda^\mu(t)) + \kappa_{\hat{f}}\mu(t) \leqslant \hat{\mathbb{L}}_\beta(x^*, y^*, \lambda^\mu(t))$, and the last inequality is established since $\alpha \geq 2$ and $\mathbb{L}_\beta(x^\mu(t), \lambda^*) - \mathbb{L}_\beta(x^*, \lambda^\mu(t)) + 2\kappa_{\hat{f}}\mu(t) \geq 0$.

The proof follows from **Theorem 17** and **Theorem 14**, which is easy to prove. Due to space limitations, we omit the proof here. ■

5. Numerical experiment

In this section, we give several numerical experiments to illustrate the effectiveness and superiority of the proposed accelerated dynamical mirror approaches. We use the **ode 15s** stiff ordinary differential equation solver in MATLAB soft to solve the dynamical approaches in all our numerical experiments.

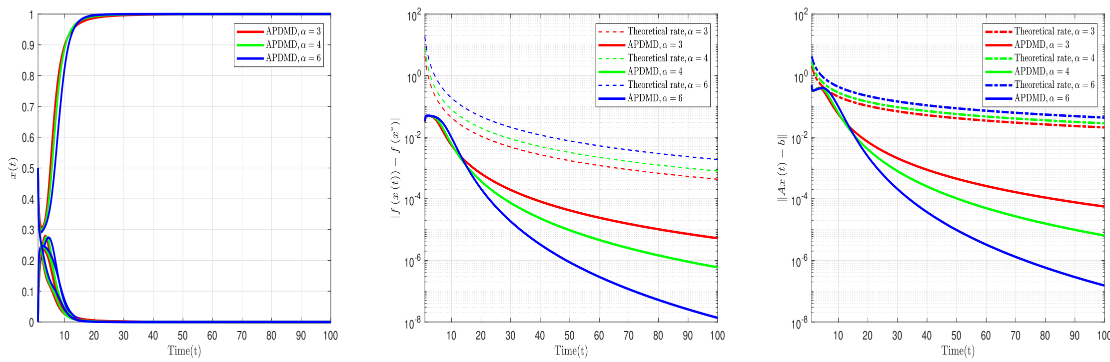

Figure 5: APDMD (14) with $\alpha = 3, 4, 6$, $\beta = 2$ and $\zeta = 0.5$ for solving the problem (99). (left) Trajectories of $x(t)$. (middle) Error of $|f(x(t)) - f(x^*)|$. (right) Error of $\|Ax(t) - b\|$.

5.1 In the smooth case

Example 2. ***Logistic regression*** *Attouch et al. (2022): Consider the problem* (1) *as follows:*

$$\begin{aligned} &\min f(x) = \log\left(1 + \exp\left(-(1,2,1,1)^T x\right)\right) \\ &\text{s.t. } Ax = b, x \in \mathcal{X}, \end{aligned} \tag{99}$$

where $A = \begin{bmatrix} 0.2, 1, 1, 2 \\ 0, 1, 0.5, 1 \end{bmatrix}$, $b = \mathrm{col}(1,1)$, $\mathcal{X} = \left\{x \in \mathbb{R}^4_+ \,|\, \sum_{i=1}^4 x_i = 1\right\}$. *The objective function in problem* (99) *is convex (but not strongly convex) and smooth, and the problem* (99) *is a very popular regularization in machine learning. Applying APDMD* (60) *with* ***Kullback-Leibler*** *divergence to address the problem* (99). *Figure 5: (left) shows the trajectories* $x(t)$ *of APDMD* (60) *versus time; Figure 5: (middle) and (right) display the error 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 3. ***Distributed Logistic regression***: *Consider a distributed problem* (12) *as follows:*

$$\begin{aligned} &\min f(x) = \sum_{i=1}^{n} \log\left(1 + \exp\left(-\left(i-1, \frac{i}{2}, i, i+1\right)^T x_i\right)\right) \\ &\text{s.t. } Lx = 0,\ x \in \prod_{i=1}^{n} \mathcal{X}_i, \end{aligned} \tag{100}$$

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* $\mathcal{X}_4 = \{x_4 \in \mathbb{R}^4 | (1, 1, 1, 1)\, x_4 \leq 4\}$. *Note that the objective function in* (100) *is convex (but not strongly convex) and smooth, which satisfies the requirement of problem* (12) *in the smooth case. Applying ADPDMD* (40) *with 4 agents (connected as a ring) to solve the problem* (100), *i.e., for agent 1, using* ***Kullback-Leibler*** *divergence; for agent 2, applying* ***Itakura-Saito*** *divergence; for agent 3, adopting projection operator of* ***ball*** *set; for agent 4, utilizing projection operator of* ***half-space*** *set. Figure 6: (left) shows the trajectories* $x(t)$ *under ADPDMD* (40) *for solving the problem* (100) *are uniformly convergent and globally asymptotically stable, i.e.,* $x_1 = x_2 = ... = x_5$; *Figure 6: (middle) and (right) display the errors of* $|f(x(t)) - f(x^*)|$ *and* $\sqrt{x(t)^T Lx(t)}$, *respectively. The numerical results show* $|f(x(t)) - f(x^*)|$ *and* $\sqrt{x(t)^T Lx(t)}$ *and* $\sqrt{x(t)^T Lx(t)}$ *all converge at the predicted rates.*

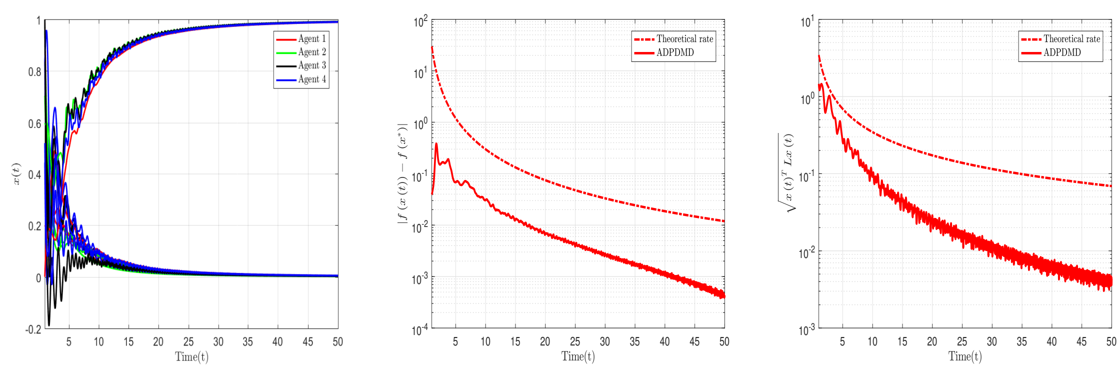

Figure 6: ADPDMD (40) with $\alpha = 3$, $\beta = 2$ and $\zeta = 0.5$ for solving the problem (100). (left) Trajectories $x(t)$. (middle) Error of $|f(x(t)) - f(x^*)|$. (right) Error of $\sqrt{x(t)^T Lx(t)}$.

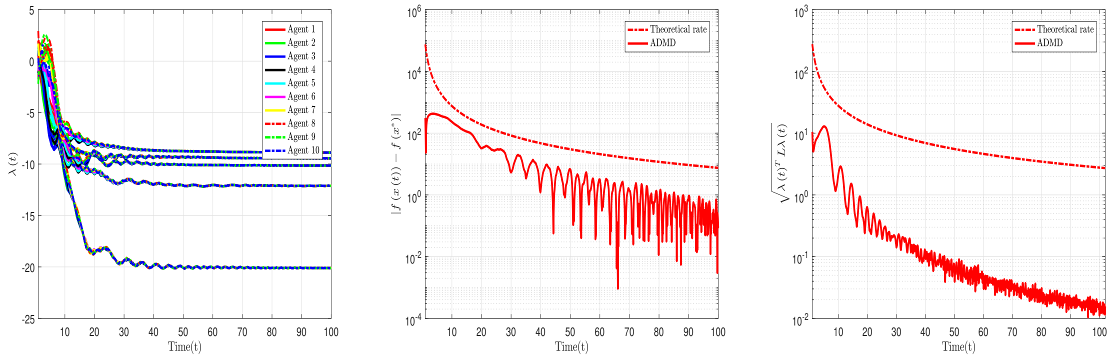

Figure 7: ADMD (50) with $\alpha = 3$, $\beta = 2$ and $\zeta = 0.5$ for the solving problem (101). (left) Trajectories of λ of 10 agents. (middle) Error of $|f(x^\mu(t)) - f(x^*)|$. (right) Error of $\sqrt{\lambda(t)^T L\lambda(t)}$.

Example 4. ***Distributed quadratic programming****: Consider a special case of DEMO (13) in the smooth case as follows:*

$$\begin{aligned} &\min f(x) = \sum_{i=1}^{n} x_i^T A_i x_i, \\ &\text{s.t. } x + Ly = 0, x \in \prod_{i=1}^{n} \mathcal{X}_i, \end{aligned} \tag{101}$$

where $\mathcal{X}_i = \{x_i \in \mathbb{R}^m | i+2 \le x_{i,j} \ll i+3, j = 1, ..., m\}, i = 1, ..., n,\ b = \mathrm{col}(7, ..., 7) \in \mathbb{R}^{nm}$

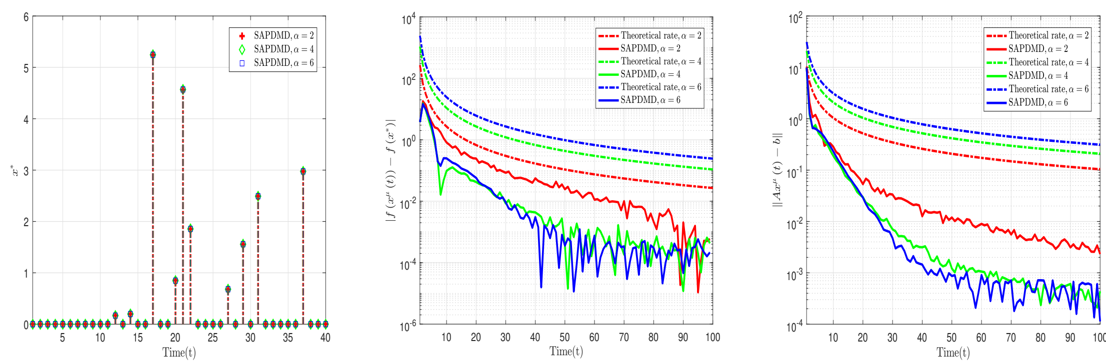

Figure 8: SAPDMD (60) with $\alpha = 2, 4, 6$, $\beta = 2$ and $\zeta = 0.5$ for solving the problem (103). (left) Reconstructed sparse signal. (middle) Error of $|f(x^\mu(t)) - f(x^*)|$. (right) Error of $\|Ax^\mu(t) - b\|$.

and $y \in \mathbb{R}^{mn}$ *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* (50) *to solve the problem* (101) *and the experimental results are shown in Figure 7. In Figure 7: (left), the dual variable λ is uniformly convergent and globally asymptotically stable, i.e., $\lambda_1 = \lambda_2 = ... = \lambda_{10}$; Figure 7: (middle and right) illustrate $|f(x^\mu(t)) - f(x^*)|$ and $\sqrt{\lambda(t)^T L\lambda(t)}$ of ADMD* (50) *are convergent at the predicted rates.*

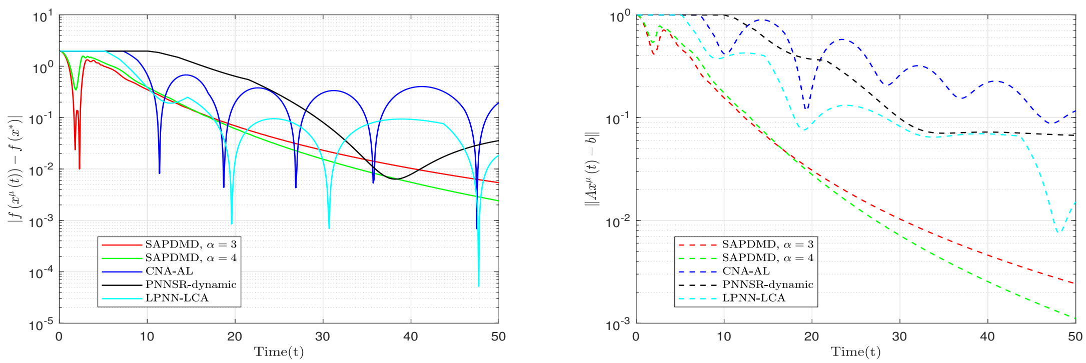

Figure 9: Convergence rates of SAPDMD (60) with $\alpha = 3, 4$, $\beta = 2$ and $\zeta = 0.5$, CNA-AL in (Zhao et al. (2021)), PNNSR in (Liu and Wang (2015)) and LPNN-LCA in (Feng et al. (2016)) for solving the problem (103). (left) Errors of $|f(x^\mu(t)) - f(x^*)|$. (right) Errors of $\|Ax^\mu(t) - b\|$.

5.2 In the nonsmooth case

Example 5. ***Nonnegative Basis Pursuit (NBP)*** *in Khajehnejad et al. (2010): Take into account a NBP as follows:*

$$\begin{aligned} &\min f(x) = \|x\|_1, \\ &\text{s.t. } Ax = b, x \in \mathcal{X}, \end{aligned} \tag{102}$$

where $\mathcal{X} = \{x \in \mathbb{R}^{128} |, x_i \geq 0, i = 1, ..., 128\}$ and $A \in \mathbb{R}^{64\times 128}$ is an orthogonal Gaussian matrix. The objective function in (102) *is convex (but not strongly convex) and nonsmooth as required. Applying the SAPDMD* (60) *and let $\psi(x) = \sum_{i=1}^n x_i \ln x_i$ to solve problem* (102)*. Figure 8 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. Note that when $x \in \mathcal{X} = \mathbb{R}^n$, the NBP* (102) *reduces to the classical Basis Pursuit (BP) problem as follows:*

$$\begin{aligned} &\min f(x) = \|x\|_1, \\ &\text{s.t. } Ax = b, \end{aligned} \tag{103}$$

where $A \in \mathbb{R}^{30\times 50}$, A is an orthogonal Gaussian matrix. We compare with SAPDMD (60) *with $\alpha = 3, 5$ to classical sparse neurodynamical approaches, CNA-AL in Zhao et al. (2021), PNNSR in (Liu and Wang (2015)) and LPNN-LCA in (Feng et al. (2016)) for solving Basis Pursuit problem* (103)*. To simulate a real sparse signal by randomly generating a signal with a sparsity of 10 (i.e., the number of non-zero elements is 10). The results of the experiment are shown in Figure 9. The numerical results of $|f(x^{\mu}(t)) - f(x^*)|$ in Figure 9: (left) and $\|Ax^{\mu}(t) - b\|$ in Figure 9: (right) illustrate that our proposed SAPDMD* (60) *with $\alpha = 3, 4$ have faster convergence rates.*

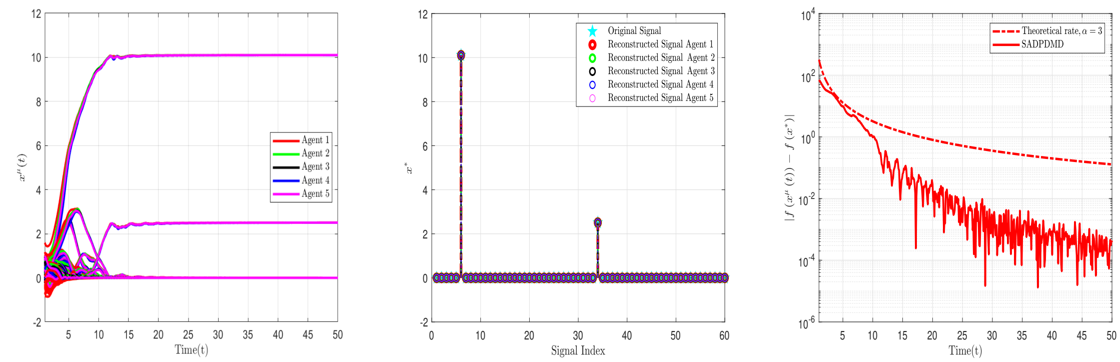

Figure 10: SADPDMD (80) with $\alpha = 3$, $\beta = 2$ and $\zeta = 0.5$ for solving the problem (104). (left) Trajectories of x^{μ} in 5 agents. (middle) Reconstructed sparse signal of 5 agents. (right) Error of $|f(x^{\mu}(t)) - f(x^*)|$.

Example 6. ***Distributed basis pursuit in row partition*** *in (Zhao et al. (2021)): Take a distributed basis pursuit with row partition of sensing matrix as follows:*

$$\begin{aligned} &\min f(x) = \sum_{i=1}^{k} \|x_i\|_1 \\ &\text{s.t. } Lx = 0, x \in \mathcal{X}, \end{aligned} \tag{104}$$

where $L = L_k \otimes I_n \in \mathbb{R}^{kn\times kn}$, $\mathcal{X} = \left\{x \in \mathbb{R}^{kn} | \hat{A}x = b, \hat{A} = \text{bldiag}\{A_{m_1\times n}, ..., A_{m_k\times n}\} \in \mathbb{R}^{m\times kn}, b \in \mathbb{R}^m\right\}$, $b \in \mathbb{R}^m$. Note that, the problem (104) *is convex (but not strongly convex) and nonsmooth and as a special case of DCCP* (12) *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* (80) *with $\nabla\psi^* = P_{\mathcal{X}}$ of affine set $\mathcal{X} = \left\{ x \in \mathbb{R}^{kn} | \hat{A}x = b \right\}$ in **Section** 2.4 to solve the problem* (104) *and all parameters comply with the requirement. The numerical results are displayed in in Figure 10. As can be seen from Figure 10: (left) that the variable x is uniformly convergent and globally asymptotically stable, i.e., $x_1^{\mu} = x_2^{\mu} = ... = x_5^{\mu}$; Figure 10: (middle) shows that the SADPDMD can efficiently solve the problem* (104) *and reconstruct the original sparse signal; the result in Figure 10: (right) is in good accordance with our theoretical results, i.e., $|f(x^{\mu}(t)) - f(x^*)|$ is convergent at the predicted rates.*

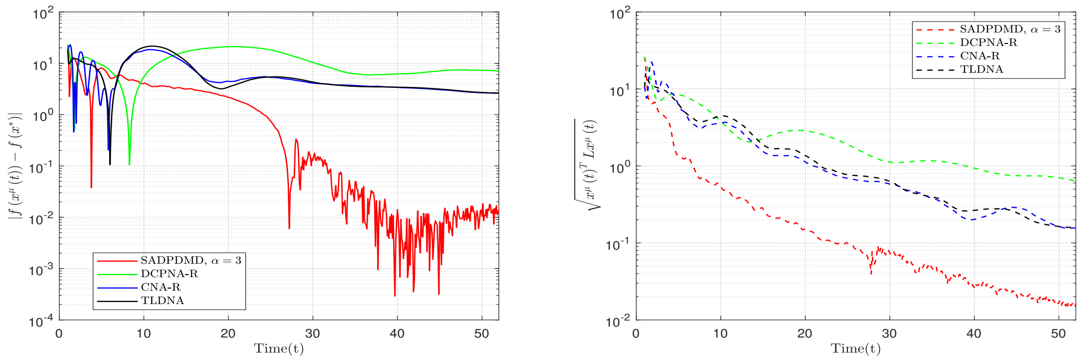

Figure 11: Convergence rates of SADPDMD (80) with $\alpha = 3$, $\beta = 2$ and $\zeta = 0.5$, XX for solving the problem (104). (left) Errors of $|f(x^{\mu}(t)) - f(x^*)|$ for the problem (103). (right) Errors of $\sqrt{x^{\mu}(t)^T L x^{\mu}(t)}$.

To demonstrate the superiority of the proposed SADPDMD (80), we compare it with $\alpha = 3$, $\beta = 2$ and $\zeta = 0.5$ to the state-of-the-art distributed sparse dynamical approaches (CNA-R in (Zhao et al. (2021)), DCPNA-R in (Zhao et al. (2022)) and TLDNA in (Xu et al. (2022)) for solving the problem (104)), and the experimental results are shown in Figure 11. The left subfigure in Figure 11 shows that SADPDMD (80) with $\alpha = 3$ has fastest convergence rate in $|f(x^{\mu}(t)) - f(x^*)|$ for solving problem (104), in addition, the right subfigure in Figure 11: (right) demonstrates that SADPDMD (80) with $\alpha = 3$ also has the fastest convergence rate in $\sqrt{x^{\mu}(t)^T L x^{\mu}(t)}$ for solving the problem (104).

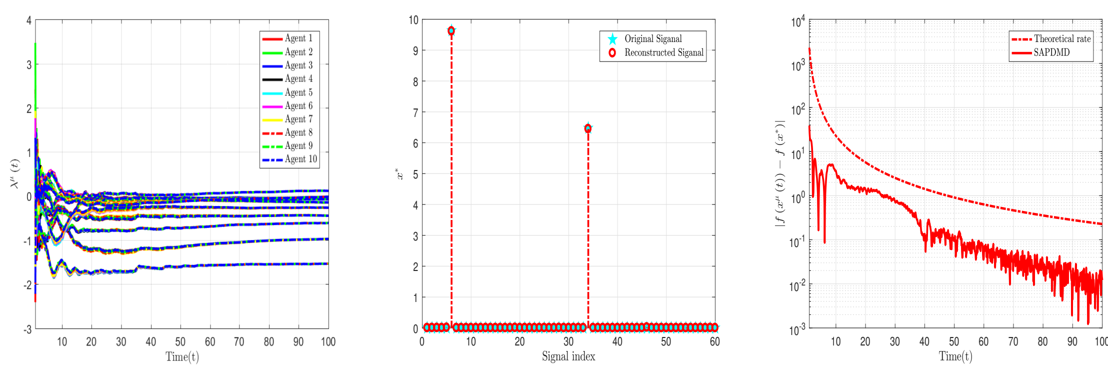

Figure 12: SADMD (89) with $\alpha = 3$, $\beta = 2$ and $\zeta = 0.5$ for solving the problem (105). (left) Trajectories of $\lambda^{\mu}(t)$ of 10 agents. (middle) Reconstructed sparse signal. (right) Error of $|f(x^{\mu}(t)) - f(x^*)|$.

Example 7. ***Distributed basis pursuit in column partition*** *in Zhao et al. (2021): Consider a distributed basis pursuit with column partition of sensing matrix as follows:*

$$\begin{aligned} &\min f(x) = \sum_{i=1}^{q} \|x_i\|_1, \\ &\text{s.t. } \bar{A}x + Ly = b, \end{aligned} \tag{105}$$

where $\bar{A} = \text{bldiag}\left\{A_{m\times n_1}, ..., A_{m\times n_q}\right\} \in \mathbb{R}^{mq\times n}, \sum_{i=1}^{q} n_q = n$, *and* $y \in \mathbb{R}^{mq}$ *is an auxiliary variable. Note that, the problem* (105) *is convex (but not strongly convex) and nonsmooth, which is a special case of DEMO* (13). *Letting* $q = 10$, *and* $A \in \mathbb{R}^{10\times 60}$, *then* $\hat{A} \in \mathbb{R}^{100\times 60}$. *Let every part* $A_{m\times n_i}, i = 1, ..., 10$ *be an agent, and 10 agents are connected as a ring. Applying the ADMD* (89) *to solve the problem* (105) *and the experimental results are shown in Figure 12. Figure 12: (left) displays the dual variable* λ *is uniformly convergent and globally asymptotically stable, i.e.,* $\lambda_1^{\mu} = \lambda_2^{\mu} = ... = \lambda_{10}^{\mu}$; *Figure 12: (middle) describes that the SADMD* (89) *is able to efficiently solve the problem* (105) *and recover the sparse signal in a distributed way; Figure 12: (right) illustrates* $|f(x^{\mu}(t)) - f(x^*)|$ *of SADMD* (89) *converges at the predicted rates.*

6. Conclusions

In this paper, we have proposed two accelerated primal-dual mirror dynamical approaches to the problem (1) in the smooth and nonsmooth cases. For the problem (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 properties of the trajectories of APDMD. We have extended the APDMD to two diustributed networks optimization problems (i.e., DCCP and DEMO) and obtained the same accelerated convergence rates as APDMD. Then, for the problem (1) with nonsmooth objective functions, 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 results of SADPDMD and SADMD when applying SAPDMD to deal with DCCP and DEMO. The obtained results can be straightly forwardly transferred to inertial (primal-dual) dynamical approaches for the problems (3), (5) and (7). How to effectively discretize the accelerated dynamical approaches proposed in this manuscript to obtain accelerated primal-dual mirror discrete-time (numerical) algorithms that match the lower bound theory is our future research direction.

Acknowledgments

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) and in part by the Fundamental Research Funds for the Central Universities (Project No. XDJK2020TY003). We would like to thank Professor Martin Jaggi and two referees for their helpful comments which helped us to improve the content of this paper.

References

Foivos Alimisis, Antonio Orvieto, Gary Bécigneul, and Aurelien Lucchi. A continuous-time perspective for modeling acceleration in riemannian optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 1297–1307. PMLR, 2020.

Mohammad S Alkousa, Alexander Vladimirovich Gasnikov, Darina Mikhailovna Dvinskikh, Dmitry A Kovalev, and Fedor Sergeevich Stonyakin. Accelerated methods for saddle-point problem. *Computational Mathematics and Mathematical Physics*, 60:1787–1809, 2020.

Zeyuan Allen-Zhu and Elad Hazan. Optimal black-box reductions between optimization objectives. volume 29, 2016.

Felipe Alvarez. On the minimizing property of a second order dissipative system in hilbert spaces. *SIAM Journal on Control and Optimization*, 38(4):1102–1119, 2000.

Hedy Attouch, Zaki Chbani, Juan Peypouquet, and Patrick Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168:123–175, 2018a.

Hedy Attouch, Zaki Chbani, and Hassan Riahi. Combining fast inertial dynamics for convex optimization with tikhonov regularization. *Journal of Mathematical Analysis and Applications*, 457(2):1065–1094, 2018b.

Hedy Attouch, Zaki Chbani, and Hassan Riahi. Rate of convergence of the nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: Control, Optimisation and Calculus of Variations*, 25:2, 2019.

Hedy Attouch, Zaki Chbani, Jalal Fadili, and Hassan Riahi. Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics. *Journal of Optimization Theory and Applications*, 193(1-3):704–736, 2022.

J-F Aujol, Ch Dossal, and Aude Rondepierre. Convergence rates of the heavy ball method for quasi-strongly convex optimization. volume 32, pages 1817–1842. SIAM, 2022.

J-F Aujol, Ch Dossal, and A Rondepierre. Convergence rates of the heavy-ball method under the łojasiewicz property. *Mathematical Programming*, 198(1):195–254, 2023.

Aurele Balavoine, Christopher J Rozell, and Justin Romberg. Convergence speed of a dynamical system for sparse recovery. *IEEE transactions on signal processing*, 61(17): 4259–4269, 2013.

Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, Joydeep Ghosh, and John Lafferty. Clustering with bregman divergences. *Journal of machine learning research*, 6(10), 2005.

HH Bauschke and PL Combettes. Convex analysis and monotone operator theory in hilbert spaces, 2011. volume 10, pages 978–1.

Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.

Pascal Bégout, Jérôme Bolte, and Mohamed Ali Jendoubi. On damped second-order gradient systems. *Journal of Differential Equations*, 259(7):3115–3143, 2015.

Wei Bian and Xiaojun Chen. Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation. *IEEE transactions on neural networks and learning systems*, 25(3):545–556, 2013.

Wei Bian and Xiaojun Chen. A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty. *SIAM Journal on Numerical Analysis*, 58(1): 858–883, 2020.

Jérôme Bolte. Sur des systèmes dynamiques dissipatifs de type gradient. applications en optimisation. 2003.

Radu Ioan Boţ and Dang-Khoa Nguyen. Improved convergence rates and trajectory convergence for primal-dual dynamical systems with vanishing damping. *Journal of Differential Equations*, 303:369–406, 2021.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Alexandre Cabot and Pierre Frankel. Asymptotics for some semilinear hyperbolic equations with non-autonomous damping. *Journal of Differential Equations*, 252(1):294–322, 2012.

Alexandre Cabot and Laetitia Paoli. Asymptotics for some vibro-impact problems with a linear dissipation term. *Journal de mathématiques pures et appliquées*, 87(3):291–323, 2007.

Xiaojun Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Mathematical programming*, 134:71–99, 2012.

Jérôme Darbon and Gabriel P Langlois. Accelerated nonlinear primal-dual hybrid gradient methods with applications to supervised machine learning. 2021.

Jelena Diakonikolas and Michael I Jordan. Generalized momentum-based methods: A hamiltonian perspective. *SIAM Journal on Optimization*, 31(1):915–944, 2021.

Mounir Elloumi, Ramzi May, and Chokri Mnasri. Asymptotic for a second order evolution equation with vanishing damping term and tikhonov regularization. *arXiv preprint arXiv:1711.11241*, 2017.

Mahyar Fazlyab, Alec Koppel, Victor M Preciado, and Alejandro Ribeiro. A variational approach to dual methods for constrained convex optimization. pages 5269–5275, 2017.

Diego Feijer and Fernando Paganini. Stability of primal–dual gradient dynamics and applications to network optimization. *Automatica*, 46(12):1974–1981, 2010.

Ruibin Feng, Chi-Sing Leung, Anthony G Constantinides, and Wen-Jun Zeng. Lagrange programming neural network for nondifferentiable optimization problems in sparse approximation. *IEEE transactions on neural networks and learning systems*, 28(10):2395–2407, 2016.

Chris Godsil and Gordon F Royle. Algebraic graph theory. volume 207. Springer Science & Business Media, 2001.

Xin He, Rong Hu, and Ya Ping Fang. Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems. *SIAM Journal on Control and Optimization*, 59(5):3278–3301, 2021.

Xin He, Rong Hu, and Ya-Ping Fang. Fast primal–dual algorithm via dynamical system for a linearly constrained convex optimization problem. *Automatica*, 146:110547, 2022.

Xin He, Rong Hu, and Ya-Ping Fang. Inertial primal-dual dynamics with damping and scaling for linearly constrained convex optimization problems. *Applicable Analysis*, 102 (15):4114–4139, 2023.

Xing He, Tingwen Huang, Junzhi Yu, Chuandong Li, and Chaojie Li. An inertial projection neural network for solving variational inequalities. *IEEE transactions on cybernetics*, 47 (3):809–814, 2016.

Xing He, Daniel WC Ho, Tingwen Huang, Junzhi Yu, Haitham Abu-Rub, and Chaojie Li. Second-order continuous-time algorithms for economic power dispatch in smart grids. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(9):1482–1492, 2017.

John J Hopfield and David W Tank. Computing with neural circuits: A model. *Science*, 233(4764):625–633, 1986.

David Alexander Hulett and Dang-Khoa Nguyen. Time rescaling of a primal-dual dynamical system with asymptotically vanishing damping. *Applied Mathematics & Optimization*, 88(2):27, 2023.

Michael Peter Kennedy and Leon O Chua. Neural networks for nonlinear programming. *IEEE Transactions on Circuits and Systems*, 35(5):554–562, 1988.

M Amin Khajehnejad, Alexandros G Dimakis, Weiyu Xu, and Babak Hassibi. Sparse recovery of nonnegative signals with minimal expansion. *IEEE Transactions on Signal Processing*, 59(1):196–208, 2010.

Nikola B Kovachki and Andrew M Stuart. Continuous time analysis of momentum methods. *The Journal of Machine Learning Research*, 22(1):760–799, 2021.

Dmitry Kovalev, Adil Salim, and Peter Richtárik. Optimal and practical algorithms for smooth and strongly convex decentralized optimization. *Advances in Neural Information Processing Systems*, 33:18342–18352, 2020.

Dmitry Kovalev, Alexander Gasnikov, and Peter Richtárik. Accelerated primal-dual gradient method for smooth and convex-concave saddle-point problems with bilinear coupling. *Advances in Neural Information Processing Systems*, 35:21725–21737, 2022.

Walid Krichene, Alexandre Bayen, and Peter L Bartlett. Accelerated mirror descent in continuous and discrete time. volume 28, 2015.

Walid Krichene, Alexandre Bayen, and Peter Bartlett. *A Lyapunov Approach to Accelerated First-Order Optimization In Continuous and Discrete Time*. PhD thesis, University of California, Berkeley, 2016.

Huan Li and Zhouchen Lin. On the complexity analysis of the primal solutions for the accelerated randomized dual coordinate ascent. *The Journal of Machine Learning Research*, 21(1):1225–1269, 2020.

Zhouchen Lin, Huan Li, and Cong Fang. *Accelerated optimization for machine learning*. Springer, 2020.

Qingshan Liu and Jun Wang. l_1-minimization algorithms for sparse signal reconstruction based on a projection neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 27(3):698–707, 2015.

Hao Luo. Accelerated primal-dual methods for linearly constrained convex optimization problems. 2021.

Ramzi May. Asymptotic for a second-order evolution equation with convex potential andvanishing damping term. *Turkish Journal of Mathematics*, 41(3):681–685, 2017.

Ramzi May, Chokri Mnasri, and Mounir Elloumi. Asymptotic for a second order evolution equation with damping and regularizing terms. *AIMS Mathematics*, 6(5):4901–4914, 2021.

Wim Michiels, Tomáš Vyhlídal, Henri Huijberts, and Henk Nijmeijer. Stabilizability and stability robustness of state derivative feedback controllers. *SIAM Journal on Control and Optimization*, 47(6):3100–3117, 2009.

Yatin Nandwani, Abhishek Pathak, and Parag Singla. A primal dual formulation for deep learning with constraints. volume 32, 2019.

Yu Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103:127–152, 2005.

Daniel O'Connor and Lieven Vandenberghe. Primal-dual decomposition by operator splitting and applications to image deblurring. *SIAM Journal on Imaging Sciences*, 7(3): 1724–1754, 2014.

Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239, 2014.

Boris T Polyak. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics*, 4(5):1–17, 1964.

R Tyrrell Rockafellar. Convex analysis. volume 11. Princeton university press, 1997.

Abdurakhmon Sadiev, Dmitry Kovalev, and Peter Richtárik. Communication acceleration of local gradient methods via an accelerated primal-dual algorithm with an inexact prox. *Advances in Neural Information Processing Systems*, 35:21777–21791, 2022.

Adil Salim, Laurent Condat, Dmitry Kovalev, and Peter Richtárik. An optimal algorithm for strongly convex minimization under affine constraints. In *International Conference on Artificial Intelligence and Statistics*, pages 4482–4498. PMLR, 2022.

Weijie Su, Stephen Boyd, and Emmanuel J Candes. A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. volume 17, pages 1–43, 2016.

Daniil Tiapkin and Alexander Gasnikov. Primal-dual stochastic mirror descent for mdps. In *International Conference on Artificial Intelligence and Statistics*, pages 9723–9740. PMLR, 2022.

Apidopoulos Vassilis, Aujol Jean-François, and Dossal Charles. The differential inclusion modeling FISTA algorithm and optimality of convergence rate in the case b ≤ 3. *SIAM Journal on Optimization*, 28(1):551–574, 2018.

Andre Wibisono, Ashia C Wilson, and Michael I Jordan. A variational perspective on accelerated methods in optimization. *proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.

Ashia C Wilson, Ben Recht, and Michael I Jordan. A lyapunov analysis of accelerated methods in optimization. *The Journal of Machine Learning Research*, 22(1):5040–5073, 2021.

Junpeng Xu, Xing He, Xin Han, and Hongsong Wen. A two-layer distributed algorithm using neurodynamic system for solving l 1-minimization. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 69(8):3490–3494, 2022.

Xianlin Zeng, Peng Yi, Yiguang Hong, and Lihua Xie. Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems. *SIAM Journal on Control and Optimization*, 56(6):3973–3993, 2018.

Xianlin Zeng, Jinlong Lei, and Jie Chen. Dynamical primal-dual nesterov accelerated method and its application to network optimization. *IEEE Transactions on Automatic Control*, 68(3):1760–1767, 2022.

Yuchen Zhang and Lin Xiao. Stochastic primal-dual coordinate method for regularized empirical risk minimization. *Journal of Machine Learning Research*, 18(84):1–42, 2017.

You Zhao, Xing He, Tingwen Huang, and Junjian Huang. Smoothing inertial projection neural network for minimization l_{p-q} in sparse signal reconstruction. *Neural Networks*, 99:31–41, 2018.

You Zhao, Xiaofeng Liao, Xing He, and Rongqiang Tang. Centralized and collective neurodynamic optimization approaches for sparse signal reconstruction via l_1-minimization. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):7488–7501, 2021.

You Zhao, Xiaofeng Liao, and Xing He. Distributed continuous and discrete time projection neurodynamic approaches for sparse recovery. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(6):1411–1426, 2022.