Abstract

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

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

©2022 * Corresponding author.
License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/ Attribution requirements are provided at http://jmlr.org/papers/v23/21-0000.html
1. Introduction

1.1 Problem statement

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

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

(1)

where

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

This problem covers many optimization problems in various applied fields such as machine learning, sparse signal reconstruction, image deblurring, resource allocation, saddle point, 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)).

1.2 Historical presentation

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

\[
\min_{x \in \mathbb{R}^n} f(x).
\]

(3)

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

\[
\text{(HBF) } \ddot{x}(t) + \eta \dot{x}(t) + \nabla f(x) = 0,
\]

(4)

where \( \eta > 0 \) is a damping parameter. Later, Alvarez (2000) studied the asymptotic behavior of (4) with a time independent parameter \( \eta \) when \( f(x) \) is convex. Bégout et al. (2015) studied some convergence properties of the HBF (4) with a constant \( \eta \) when \( f(x) \) is a nonconvex function. Aujol et al. (2022, 2020) investigated the convergence rate of HBF dynamical system and its corresponding discrete algorithm with a fixed parameter \( \eta \) under the condition that \( f(x) \) satisfies quasi-strongly convex and Lojasiewicz properties, respectively. Su et al. (2016) first revealed that the HBF (4) with \( \eta = \frac{\alpha}{T}, \alpha \geq 3 \) can be regarded as the continuous-time limit of the Nesterov’s accelerated gradient algorithm and it has an accelerated convergence rate, i.e., \( f(x(t)) - \min f(x) = O\left(\frac{1}{t^2}\right) \). Elloumi et al. (2017) further improved the convergence rate of it to \( f(x(t)) - \min f(x) = o\left(\frac{1}{t^2}\right) \). When \( \eta = \frac{\alpha}{T}, 0 < \alpha \leq 3 \), the convergence rate \( f(x(t)) - \min f(x) = O\left(t^{-\frac{3\alpha}{2\gamma}}\right) \) was estimated by Attouch et al. (2019) and Apidopoulos et al. (2018) for smooth and nonsmooth convex functions \( f(x) \). Cabot and Frankel (2012) studied the asymptotic behavior of the HBF (4) when \( \eta = \frac{\alpha}{T^\gamma}, 0 < \gamma < 1 \). 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 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.,

\[
\min_{x \in \mathbb{R}^n} f(x), \quad \text{s.t. } x \in \mathcal{X} \subset \mathbb{R}^n,
\]

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, two smoothing inertial hydrodynamic approaches were proposed to solve constrained nonconvex \( L_{p-q} \) 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 problem (5) as follows:

\[
\begin{align*}
\dot{X} &= \gamma t (\nabla \psi^* (Z) - X), \\
\dot{Z} &= -t \nabla f (X), \\
X(0) &= x_0, Z(0) = z_0, \text{ with } \nabla \psi^* (z_0) = x_0,
\end{align*}
\]

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

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

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

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 between the Lagrangian function and its optimal value is \( O \left( \frac{1}{t^2} \right) \). The corresponding inertial dynamical approach is then extended to solve the distributed optimization problems. Later, Bot 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/perturbations for separable convex optimization problems with an affine constraint. Atouch et al. (2021) 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 guarantees.

However, many practical problems in machine learning, sparse signal reconstruction, image deblurring, resource allocation and so on, not only have affine constraints, but also have set constraints, as in problem (1). The problem (1) can be regarded as a generalization for the problems (3), (5), (7), (12) and (13). In recent years, some first order dynamical
approaches based on primal-dual framework and projection were proposed by Liu et al. (2017); Qu and Li (2019); Yi et al. (2016); Zeng et al. (2018) to solve the problem (1) only with convergence analysis or ergodic $O \left( \frac{1}{t} \right)$ convergence rate. To the best of our knowledge, accelerated (inertial) dynamical approaches based on primal-dual framework with non-ergodic $O \left( \frac{1}{t^2} \right)$ convergence rates are rarely involved. It is worth noting that the inertial dynamical approaches mentioned above were used for solving the problems (3), (5) and (7) and they cannot be used directly for the problem (1).

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 Lagrangian and objective functions’ gaps. We provide an interpretation of the APDMD from different perspectives (i.e., neurodynamic approach, Hamilton’s system, game theory and control theory). Moreover, based on the properties of conjugate function and Cauchy-Lipschitz-Picard theorem, the feasibility (i.e., ensuring that the trajectories of solutions always satisfy the set constraint), existence and uniqueness of the global solution of APDMD are obtained. Last, applying the APDMD to DCCP and EDMO leads to two distributed APDMDs (ADPDMD and ADMD) with accelerated convergence guarantees.

- For the problem (1) 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 condition. Last, applying the SAPDMD to DCCP and EDMO leads to two distributed SAPDMDs (SADPDMD and SADMD) with accelerated convergence guarantees.

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

\[
g_f (x) = \{ p \in \mathbb{R}^n | f (z) - f (x) \geq p^T (z - x), \forall x, z \in \mathcal{X} \},
\]

and the element \( \partial f (x) \) of \( g_f (x) \) is called subgradient of \( f (x) \). In addition, if \( f \) is smooth, the subgradient \( \partial f (x) \) reduces to gradient \( \nabla f (x) \).

2.2 Saddle point Theorem

The (augmented) Lagrangian \( \mathcal{L}_\beta : \mathcal{X} \times \mathbb{R}^m \to \mathbb{R} \) with respect to the problem (1) with \( \beta \geq 0 \) is defined by

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

(8)

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

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

(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 \( \psi \) is

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

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

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

from the Fenchel’s duality theorem.

Assume that \( \psi \) and \( \psi^* \) are proper and convex functions, then they’re subdifferentiable, i.e., \( \partial \psi (x), \partial \psi^* (u) \) exist in the relative interior of their domains \( \mathcal{X}, \mathcal{X}^* \), respectively Rockafellar (1970). In addition, using the condition that \( \psi^{**} (u) = \psi (x) \), one has

\[
\partial \psi^* (u) + \partial \psi (x) = u^T x \iff u \in \partial \psi (x) \iff x \in \partial \psi^* (u),
\]

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

**Definition 1.** A convex function \( \psi \) is cofinite if its epigraph does not consist of any non-vertical half-line.
Definition 2. A convex function \( \psi \) is essentially strictly convex if it is strictly convex and subdifferentiable on all convex subsets.

Lemma 3. If \( \psi \) and \( \psi^* \) are proper, convex, and closed, so they are inverses of each other, then \( \psi^* \) is finite and differentiable on \( X^* \) if and only if \( \psi \) is essentially strictly convex and cofinite.

Lemma 4. Bregman divergence: The distance between \( h \) at \( y \in X \) and its first-order Taylor series approximation at \( z \in X \) is given by:

\[
D_h(y, z) = h(y) - h(z) - \nabla h(z)^T (y - z), \forall y, z \in X,
\]

which is nonnegative if \( h \) is convex and is an approximation to the Hessian metric, i.e.,

\[
D_h(y, z) \approx \frac{1}{2} (y - z)^T \nabla^2 h(x) (y - x), \forall y, z \in X \text{ when } y \text{ is close to } z.
\]

2.4 Projection operators

Define the projection operator of a closed and convex set \( X \) at \( x \) by \( P_X(u) = \arg \min_{z \in X} \|u - z\| \).

A basic property of the projection operator \( P_X \) is

\[
(P_X(x) - z)^T (P_X(x) - x) \leq 0, \forall x \in \mathbb{R}^n, z \in X.
\]

Lemma 5. 1): If \( X \) is a box set, i.e. \( X = \{ u \in \mathbb{R}^n \mid u_i \leq u \leq \bar{u}, i = 1, \ldots, n \} \), then, for any \( i \), we have

\[
P_X(u_i) = \begin{cases} 
   u_i, & \text{if } u_i < u_i, \\
   u, & \text{if } u_i \leq u_i \leq \bar{u}_i, \\
   \bar{u}_i, & \text{if } u_i > \bar{u}_i.
\end{cases}
\]

2): If \( X \) is a Sphere set, i.e., \( X = \{ u \in \mathbb{R}^n \mid \| u - z \| \leq r, z \in \mathbb{R}^n, r > 0 \} \), then, the projection operator is

\[
P_X(u) = \begin{cases} 
   u, & \text{if } \| u - z \| \leq r, \\
   z + \frac{r(u - z)}{\|u - z\|}, & \text{if } \| u - z \| > r.
\end{cases}
\]

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

4): If \( X \) is a half-space set, i.e., \( X = \{ u \in \mathbb{R}^n \mid a^Tu \leq b, a \neq 0 \} \), then, the projection operator is

\[
P_X(u) = \begin{cases} 
   u + \frac{b - a^T u}{\|a\|^2} a, & \text{if } a^T u > b, \\
   u, & \text{if } a^T u \leq b.
\end{cases}
\]

For more information regarding the projection operators, (please refer to Bauschke and Combettes (2011); Parikh and Boyd (2014)).

2.5 Graph Theory

An undirected communication topology graph is a triplet \( G = (V, E, A) \) with node set \( V = \{v_1, v_2, \ldots, v_n\} \), edge set \( E \subseteq V \times V \) and connection matrix \( A = \{a_{ij}\}_{n \times n} \) with nonnegative elements \( a_{ij} = a_{ji} \geq 0 \) if \((i, j) \in E\), and \( a_{ij} = a_{ji} = 0 \) 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 $i, i_1, \ldots, i_s, j$ denote different agents. Let $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$ be an agent $i$’s neighbor set. The undirected graph $G$ is connected if there exists a path between any pair of distinct nodes $v_i$ and $v_j$ ($i, j = 1, 2, \ldots, n$) (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) $\lim_{x \to w, \mu \to 0} \hat{f} (x, \mu) = f (w), \forall x \in \mathcal{X}$; 

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

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

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

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

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

$$\left| \hat{f} (x, u_1) - \hat{f} (x, u_2) \right| \leq \kappa_{\hat{f}} |u_1 - u_2|;$$

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

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

The smoothing function satisfying the above conditions in (6) enjoys the following properties (see Bian and Chen (2014)):

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

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 $l_{\Psi}$. Let $\hat{\Phi}$ be a smoothing function of $\Phi$, then $\Psi (\hat{\Phi})$ is a smoothing function of $\Psi (\Phi)$ with $\kappa_{\Psi (\hat{\Phi})} = l_{\Psi} \kappa_{\hat{\Phi}}$.

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 $\Phi$, 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 (6). A smoothing function for the \( g(s) = \max \{0, s\} \) is given by

\[
\hat{g}(s, \mu) = \begin{cases} 
\max \{0, s\}, & \text{if } |s| > \mu, \\
\frac{(s+\mu)^2}{4\mu}, & \text{if } |s| \leq \mu,
\end{cases}
\] (10)

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

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

\[
\begin{align*}
\max \{s, x\} &= s + \max \{s - x, 0\}, \\
\min \{s, x\} &= s - \max \{s - x, 0\}, \\
\text{mid} \{s, x, w\} &= \min \{\max \{x, s\}, w\}, \\
\theta(s) &= |s| = \max \{s, 0\} + \max \{-s, 0\},
\end{align*}
\]

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

\[
\hat{\theta}(s, \mu) = \begin{cases} 
|s|, & \text{if } |s| > \frac{\mu}{2}, \\
\frac{s^2}{\mu} + \frac{\mu}{4}, & \text{if } |s| \leq \frac{\mu}{2},
\end{cases}
\] (11)

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

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

Figure 1: (left) The smoothing function \( \hat{g}(s, \mu) \) with different \( \mu \). (right) The smoothing function \( \hat{\theta}(s, \mu) \) with different \( \mu \).
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. $\text{col}(x_1, ..., x_n) = (x_1^T, ..., x_n^T)^T$, $\|\cdot\|$ is the Euclidean norm, $\|\cdot\|_1$ is the 1-norm. $L^{1}_{\text{loc}}([t_0, +\infty))$ represents the local Lebesgue integrable functions on $[t_0, +\infty)$. Let $A_i \in \mathbb{R}^{p_i \times q_i}$ ($p_i, q_i > 0$), $i = 1, ..., n$, then we define $\bar{A} = \text{bddiag}\{A_1, ..., A_n\} \in \mathbb{R}^{\sum_{i=1}^{n} p_i \times \sum_{i=1}^{n} q_i}$, which means a block diagonal matrix. For $x_i \in \mathbb{R}^m$, $i = 1, ..., n$, we use $\text{col}(x_1, ..., x_n) \in \mathbb{R}^{mn}$ to denote an $mn$ column vector.

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 (31) and ADMD (36), 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**

1. The objective function $f(x)$ is smooth and convex on an open set containing $\mathcal{X}$, where $\mathcal{X}$ is a closed and convex set;
2. The function $\psi$ is proper, essentially strictly convex and cofinite (see (2.3));
3. (Slater’s condition) There exists a vector $x \in \text{int}(\mathcal{X})$ 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 is to cooperate for seeking the minimum of the following problem:

$$
\min_{x \in \mathbb{R}^{nm}} f(x) = \sum_{i=1}^{n} f_i(x_i),
$$

s.t. $L x = 0$ (i.e., $x_i = x_j$, $i = 1, ..., n$, $x_i \in \mathcal{X}_i \subset \mathbb{R}^m$),

where $x = \text{col}(x_1, ..., x_n) \in \mathbb{R}^{nm}$, and $\mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i \subset \mathbb{R}^{nm}$.

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

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

(13)
where \( A = [A_1, ..., A_n] \in \mathbb{R}^{m \times \sum_{i=1}^n p_i} \) and \( A_i \in \mathbb{R}^{m \times p_i} \).

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

**Assume 1:**
1. The objective function \( f(x) = \sum_{i=1}^m f_i(x_i) \) and for all \( i = 1, ..., n, \) \( f_i(x_i) \) is smooth and convex on an open set containing \( \mathcal{X}_i \), and \( \mathcal{X}_i \) is a closed and convex set;
2. For all \( i = 1, ..., n, \) \( \psi_i \) is proper, essentially strictly convex and cofinite (see (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:

\[
\begin{align*}
\dot{x}(t) &= \frac{\beta}{\alpha} (\nabla \psi^*(u(t)) - x(t)), \\
\dot{u}(t) &= -\frac{1}{\alpha} (\nabla f(x(t)) + \beta A^T(Ax(t) - b) + A^Tv(t)), \\
\dot{\lambda}(t) &= \frac{\alpha}{\beta} (v(t) - \lambda(t)), \\
\dot{v}(t) &= \frac{T}{\beta} (A\nabla \psi^*(u(t)) - b). \\
x(t_0) = x_0, u(t_0) = u_0 \text{ with } \nabla \psi^*(u_0) = x_0 \in \mathcal{X}, \lambda(t_0) = \lambda_0, v(t_0) = v_0,
\end{align*}
\]

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

Note that the APDMD (14) can be equivalently transformed into the following second-order dynamical approach:

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

where the Hessian term \( \nabla^2 \psi^* \left( \nabla \psi \left( x(t) + \frac{T}{\alpha} \dot{x}(t) \right) \right) \) is nonlinear transformation. It applies to \( \nabla f(x(t)) + \beta A^T(Ax(t) - b) + A^T \left( \lambda(t) + \frac{T}{\alpha} \dot{\lambda}(t) \right) \) to guarantee the trajectory of \( x \) is inside the feasible set \( \mathcal{X} \) in the intuitive understanding (the rigorous proof process in (8)). The form in (15) 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 \( \lambda \).

### 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 be implemented by analog circuit like the Hopfield neural network in Hopfield and
Figure 2: (left) Illustration of the APDMD (14). (right) The feasibility of \( x(t) \) to APDMD (14).

Figure 3: Circuit architecture diagram of APDMD (14) with

1. \( \frac{1}{\alpha} \dot{x} + x = \nabla \psi^*(u) \),
2. \( \dot{u} = -\frac{1}{\alpha} (\nabla f(x) + \beta A^T (Ax - b) + A^T v) \),
3. \( \frac{1}{\alpha} \dot{\lambda} + \lambda = v \),
4. \( \dot{v} = \frac{1}{\alpha} (A \nabla \psi^*(u) - b) \).
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 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

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

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

The corresponding continuous-time dynamics of $H_1$ is

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

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

$$H_2 (\lambda (t) , v (t) , \eta (t)) = \frac{1}{2} \| v (t) \|^2 - \Gamma (\eta (t)) \left( \frac{\lambda (t) }{\eta (t)} \right)^T (A u (t) - b),$$

where $\lambda (t) = \eta (t) \lambda (t)$.

The corresponding continuous-time dynamics of $H_2$ is

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

From the above conclusions, the APDMD (14) can be obtained directly.
• **Game theoretic standpoint**  Attouch et al. (2021). 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., \( x(t) + \frac{\varphi}{T} \dot{x}(t) \). Conversely, for player \( x(t) \), it takes account of the anticipated position of the player \( \lambda(t) \), i.e., \( \lambda(t) + \frac{\varphi}{T} \dot{x}(t) \).

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

\[
\dot{\chi}(t) + \frac{\alpha + 1}{T} \chi(t) = \Upsilon(t, \chi(t), \dot{\chi}(t)),
\]

with an operator \( \Upsilon \), i.e., a feedback control term which takes the constraint into account. It is a function of the state \( \chi(t) \), its derivative \( \dot{\chi}(t) \) and time \( t \). For a comprehensive understanding of state derivative feedback, the readers can consult reference ?.

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

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

since \( d(\nabla \psi^*(u_0)) \in \mathcal{X} \). 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 \( t_0 > 0 \). \((x,u,\lambda,v)\) 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) The APDMD holds for almost every \( t \geq t_0 \),

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

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

(a) There exists \( x : [t_0, T) \to \mathbb{R}^n \) an integrable function, such that
\[
x(t) = x(t_0) + \int_{t_0}^{t} x(s) \, ds, \forall t \in [t_0, T];
\]

(b) \( x \) is a continuous function and its distributional derivative is Lebesgue integrable on the interval \([t_0, T]\);

(c) For every \( \varepsilon > 0 \), there exists \( \pi > 0 \), such that for every finite family \( I_k = (a_k, b_k) \) from \([t_0, T]\), the following implication is valid:
\[
\left[ I_k \cap I_j = \emptyset \text{ and } \sum_k |b_k - a_k| < \pi \right] \Rightarrow \left[ \sum_k \|x(b_k) - x(a_k)\| < \varepsilon \right].
\]

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:

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

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

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 \( I(t) \)-Lipschitz continuous and \( I(\cdot) \in L^1_{\text{loc}}([t_0, +\infty)) \).

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

The proof of (I). Let \( t \in [t_0, +\infty) \) be fixed and utilize the Lipschitz continuous of \( \nabla \psi^*, \nabla f \) and \( \|X + Y\|^2 \leq 2 \|X\|^2 + 2 \|Y\|^2 \) for any vectors \( X \) and \( Y \), then, for any \( Y \) and \( \dot{Y} \), we have
\[
\|F(t, Y) - F(t, \dot{Y})\| \leq \left( \frac{t^2}{\alpha^2} \left( 4\beta^2 \delta_{\text{max}} (A^T A)^2 + 4\beta^2 + \delta_{\text{max}} (A^T A) I_{\psi^*}^2 + 2\delta_{\text{max}} (A^T A) \right) + \frac{\alpha^2 (6 + 2 l_{\psi^*}^2)}{t^2} \right)^{\frac{1}{2}} \|Y - \dot{Y}\|,
\]

Y. Zhao, X. F. Liao, X. He and C. Li
where $\delta_{\max}(A)$ is the maximum singular value of matrix $A$ and the inequality holds from the Lipschitz continuous properties of $\nabla f$ and $\nabla \psi^*$ that have Lipschitz constants $l_f$ and $l_{\psi^*}$. Let $I(t)$ be \( \left( t^2 \left( 4\beta^2 \delta_{\max}(A^T A)^2 + 4t^2 \delta_{\max}(A^T A)^2 \alpha^2 + 2\delta_{\max}(A^T A) \right) + \frac{\alpha^2}{t^2} (6 + 2t^2 \delta_{\max}) \right) \), one has
\[
\left\| F(t, Y) - F(t, \hat{Y}) \right\| \leq I(t) \left\| Y - \hat{Y} \right\| .
\]
Note that $I(t)$ is continuous on $[t_0, +\infty)$. Hence $I(\cdot)$ is integrable on $[t_0, T]$ for all $0 < t_0 < T < +\infty$.

The proof of (II). For arbitrary $Y \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m$ and $0 < t_0 < T$, we have
\[
\int_{t_0}^{T} \left\| F(t, Y(t)) \right\| dt
= \int_{t_0}^{T} \left( \frac{\alpha^2}{t^2} \left\| \nabla \psi^* (u(t)) - x(t) \right\|^2 + \frac{\alpha^2}{t^2} \left\| v(t) - \lambda(t) \right\|^2 + \frac{t^2}{\alpha^2} \left\| A (\nabla \psi^* (u(t)) - x^*) \right\|^2 + \frac{t^2}{\alpha^2} \left\| \nabla f (x(t)) + \beta A^T A (x(t) - x^*) - A^T v(t) \right\|^2 \right) dt
\leq \int_{t_0}^{T} \left( \frac{2\alpha^2}{t^2} \left\| \nabla \psi^* (u(t)) \right\|^2 + \frac{4\beta^2 \delta_{\max}(A^T A)^2}{\alpha^2} \left\| x(t) \right\|^2 + \frac{2\alpha^2}{t^2} \left\| v(t) \right\|^2 + \frac{2\alpha^2}{t^2} \left\| \lambda(t) \right\|^2 + \frac{2\alpha^2}{t^2} \left\| x(t) \right\|^2 + \frac{2\alpha^2}{t^2} \left\| x(t) \right\|^2 + \frac{2t^2 \delta_{\max}(A^T A)}{\alpha^2} \left\| \nabla \psi^* (u(t)) \right\|^2 + \frac{2t^2 \delta_{\max}(A^T A)}{\alpha^2} \left\| x^* \right\|^2 + \frac{4t^2 \delta_{\max}(A^T A)}{\alpha^2} \left\| \nabla f (x(t)) \right\|^2 + \frac{4t^2 \delta_{\max}(A^T A)}{\alpha^2} \left\| v(t) \right\|^2 + \frac{4t^2 \delta_{\max}(A^T A)}{\alpha^2} \left\| x^* \right\|^2 \right) dt
\leq \sqrt{\left\| \nabla \psi^* (u) \right\|^2 + \left\| x \right\|^2 + \left\| \lambda \right\|^2 + \left\| v \right\|^2 + \left\| x^* \right\|^2 + \left\| \nabla f (x(t)) \right\|^2}
\times \left( 8\alpha^2 \frac{t^2}{t^2} + \frac{2\alpha^2}{t^2} \left( 4\delta_{\max}(A^T A) + 4\beta^2 \delta_{\max}(A^T A)^2 + 2 \right) \right),
\]
and the conclusion holds by the continuity of the following function
\[
t \to \left( 8\alpha^2 \frac{t^2}{t^2} + \frac{2\alpha^2}{t^2} \left( 4\delta_{\max}(A^T A) + 4\beta^2 \delta_{\max}(A^T A)^2 + 2 \right) \right)^{1/2} \text{ on } [t_0, T].
\]

In view of the above statements (I) and (II), the existence and uniqueness of a strong global solution for the dynamical system (16) 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 \((x^*, \lambda^*)\) be the solution trajectory and optimal solution for APDMD (15) and problem (1), respectively. Then for any \((x(t_0), u(t_0), \lambda(t_0), v(t_0)) \in \mathcal{X} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m\) with \(\alpha \geq 2\), the following statements are true:

(I): The trajectories \(x(t)\) and \(\lambda(t)\) of APDMD (15) are bounded for any \(t \geq t_0 > 0\).

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

\[
L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t)) \leq \frac{\alpha^2 V(t_0)}{t^2},
\]

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

\[
\int_0^{+\infty} t (L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t))) dt < +\infty,
\]

\[
\int_{t_0}^{+\infty} t \|Ax(t) - b\|^2 dt < +\infty.
\]

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

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

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

Proof Consider a Lyapunov function \(V : [t_0, T] \rightarrow [0, +\infty)\) which is given by

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

where

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

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

Note that the Lyapunov function \(V_1(t)\) is positive and radially unbounded since \(L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t)) \geq 0, \forall (x(t), \lambda(t)) \in \mathcal{X} \times \mathbb{R}^m\) in (9), \(\alpha \geq 2\) and \(t \geq t_0 > 0\). For \(V_2(t)\) and \(V_3(t)\), since \(\psi^* (\cdot)\) and \(h(\cdot) = \frac{1}{2} \|\cdot\|^2\) are convex (Assumption 3.1 holds), we get \(D_{\psi^*} (u(t), u^*) \geq 0, D_h (v(t), \lambda^*) \geq 0\), moreover, \(\|u(t)\| \rightarrow +\infty, V_2(t) \rightarrow +\infty\) and \(\|v(t)\| \rightarrow +\infty, V_3(t) \rightarrow +\infty\). Thus, both \(V_2(t)\) and \(V_3(t)\) are positive and radially unbounded.
The derivatives of $V_1(t)$, $V_2(t)$ and $V_3(t)$ along the trajectory of APDMD with $x(t) \in \mathcal{X}$ are given by

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

(20)

$$
\dot{V}_2(t) = (\nabla \psi^*(u(t)) - \nabla \psi^*(u^*))^T \dot{u}(t)
= -\frac{t}{\alpha} (\nabla \psi^*(u(t)) - \nabla \psi^*(u^*))^T
\times (\nabla f(x(t)) + \beta A^T (Ax(t) - b) + A^T v(t))
= -\frac{t}{\alpha} (\nabla \psi^*(u(t)) - \nabla \psi^*(u^*))^T
\times (\nabla f(x(t)) + \beta A^T A(x(t) - x^*) + A^T \lambda^*)
- \frac{t}{\alpha} (\nabla \psi^*(u(t)) - \nabla \psi^*(u^*))^T (A^T v(t) - A^T \lambda^*),
$$

(21)

$$
\dot{V}_3(t) = (v(t) - \lambda^*)^T \dot{v} = \frac{t}{\alpha} (v(t) - \lambda^*)^T A (\nabla \psi^*(u(t)) - x^*).
$$

(22)

Adding (20), (21) and (22) together, and using $Ax = b$, we have

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

(23)

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 $\lambda^*$, the third equation holds because of $\mathcal{L}_\beta(x^*, \lambda(t)) = \mathcal{L}_\beta(x^*, \lambda^*)$ and the last inequality is established since $\alpha \geq 2$ and $\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) \geq 0$. 

17
Since the Lyapunov function $V(t)$ is positive and radially unbounded with any $t \geq t_0 > 0$. This implies that the trajectories of $x(t)$ and $\lambda(t)$ are bounded for any $t \geq t_0 > 0$. The proof of (I) therefore completed.

(II): From (23), we have that $\dot{V}(t) \leq 0$, i.e., $V(t)$ is nonincreasing on $[t_0, +\infty)$. Thus, for any $t \geq t_0 > 0$ it yields to $\frac{t^2}{\beta} (\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t))) = V_1(t) \leq V(t) \leq V(t_0)$, i.e., $\mathcal{L}_\beta(x, \lambda^*) - \mathcal{L}_\beta(x^*, \lambda) \leq \frac{\alpha^2 V(t_0)}{t_0}$, thus the condition (17a) holds. In addition, the (17a) further implies $\|Ax(t) - b\|^2 \leq \frac{2\alpha^2 V(t_0)}{\beta t_0}$, i.e., (17b) is satisfied.

Since $\dot{V}(t) \leq -t\frac{\alpha - 2}{\alpha^2} (\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t))) - \frac{\beta}{2\alpha} \|Ax(t) - b\|^2$, thus, integrating the above inequality from $t_0$ to $+\infty$ and rearranging it, we have

$$\frac{\alpha - 2}{\alpha^2} \int_{t_0}^{+\infty} \frac{\alpha - 2}{t} \mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t))) dt + \frac{\beta}{2\alpha} \int_{t_0}^{+\infty} t \|Ax(t) - b\|^2 dt$$

$$= V(t_0) - V(\infty) \leq V(t_0) < +\infty,$$

and it combines with $\mathcal{L}_\beta(x(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda(t)) \geq 0, \forall(t) \in \mathcal{X}$ leads to

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

i.e., the condition (17c) holds. Moreover, (17c) combines with $\|Ax(t) - b\|^2 \geq 0$ implies $\int_{t_0}^{+\infty} t \|Ax(t) - b\|^2 dt < +\infty$, i.e., the (17d) holds.

(III): The equation $\lambda^* = \begin{cases} 0, & \text{if } Ax(t) - b = 0, \\ \frac{Ax(t) - b}{\|Ax(t) - b\|}, & \text{if } Ax(t) - b \neq 0, \end{cases}$ and (17a) yields $0 \leq f(x(t)) - f(x^*) + \|Ax(t) - b\| + \frac{\beta}{2} \|Ax(t) - b\|^2 \leq \frac{\alpha^2 V(t_0)}{t_0}$.

Combining the (24), $\|Ax(t) - b\| \geq 0$ and $f(x(t)) - f(x^*) \leq \frac{\alpha^2 V(t_0)}{t^2}$, we deduce that

$$-\frac{\alpha\sqrt{V(t_0)}}{t} \leq f(x(t)) - f(x^*) \leq \frac{\alpha^2 V(t_0)}{t^2}$$

with any $t \geq t_0 > 0$. Finally, the conclusion (III) holds.

3.5 Examples of the APDMD

In this subsection, we will illustrate some examples of the APDMD (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

3.6 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
3.7 Examples of the APDMD

**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 \( \nabla \psi \left( x(t) + \frac{\alpha}{t} \dot{x}(t) \right) = u(t), \) \( \psi^* (u(t)) = \frac{1}{2} \left\| u(t) \right\|^2, \) \( \nabla \psi^* (u(t)) = u(t), \) \( \nabla^2 \psi^* \left( \nabla \psi \left( x(t) + \frac{\alpha}{t} \dot{x}(t) \right) \right) \equiv I_n \). The APDMD (15) reduces to the classical accelerated primal-dual dynamical approaches He et al. (2021); Zeng et al. (2022); Boţ and Nguyen (2021); Attouch et al. (2021).

\[
\begin{cases}
\dot{x}(t) + \frac{\alpha + 1}{t} \dot{x}(t) + \left( \nabla f \left( x(t) \right) + \beta A^T (Ax(t) - b) + AT \left( \lambda(t) + \frac{\alpha}{t} \dot{\lambda}(t) \right) \right) = 0, \\
\dot{\lambda}(t) + \frac{\alpha + 1}{t} \dot{\lambda}(t) - A \left( x(t) + \frac{\alpha}{t} \dot{x}(t) \right) + b = 0, \\
x(t_0) = x_0 \in \mathcal{X}, \dot{x}(t_0) = \dot{x}_0, \lambda(t_0) = \lambda_0, \dot{\lambda}(t_0) = \dot{\lambda}_0.
\end{cases}
\]

**Case 2:** 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 \( \nabla \psi \left( x(t) \right) = \text{col} \left( 1 + \ln x_1, ..., 1 + \ln x_n \right) \) and \( \psi^* (u) = \sum_{i=1}^n e^{u_i - 1} \), \( \nabla^2 \psi^* (u(t)) = \text{diag} \left( e^{u_1 - 1}, ..., e^{u_n - 1} \right) \). Thus, for all \( x \in \mathbb{R}^n_+ \), we have

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

Correspondingly, APDMD (15) is turned into

\[
\begin{cases}
\dot{x}(t) + \frac{\alpha + 1}{t} \dot{x}(t) + \text{diag} \left( \frac{t}{\alpha} \dot{x}(t) \right) \left( \nabla f \left( x(t) \right) \right) + \beta A^T (Ax(t) - b) + AT \left( \lambda(t) + \frac{\alpha}{t} \dot{\lambda}(t) \right) = 0, \\
\dot{\lambda}(t) + \frac{\alpha + 1}{t} \dot{\lambda}(t) - \left( x(t) + \frac{\alpha}{t} \dot{x}(t) \right) + b = 0, \\
x(t_0) = x_0 \in \mathcal{X}, \dot{x}(t_0) = \dot{x}_0, \lambda(t_0) = \lambda_0, \dot{\lambda}(t_0) = \dot{\lambda}_0.
\end{cases}
\]

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

\[
\nabla^2 \psi^* \left( \nabla \psi \left( x(t) + \frac{t}{\alpha} \dot{x}(t) \right) \right) = \text{diag} \left( \left( x_1(t) + \frac{t}{\alpha} \dot{x}_1(t) \right)^2, ..., \left( x_n(t) + \frac{t}{\alpha} \dot{x}_n(t) \right)^2 \right).
\]
**Case 3:** Beck and Teboulle (2003) If $\mathcal{X}$ is a unit simplex set, i.e., $\mathcal{X} = \Delta = \{ x \in \mathbb{R}_+^n | \sum_{i=1}^n x_i = 1 \}$. Considering a distance-generating function $\psi (x) = \sum_{i=1}^n x_i \ln x_i + I_X (x)$, where $\sum_{i=1}^n x_i \ln x_i$ is the negative entropy function and $I_X (x)$ is the indicator function on $\mathcal{X}$ and its Bregman divergence is $D_\psi (x,y) = \sum_{i=1}^n x_i \ln \frac{x_i}{y_i}$ between the vectors $x$, $y$ (called Kullback-Leibler divergence). Then, we have

$$\nabla \psi (x) = \text{col} (1 + \ln x_1, \ldots , 1 + \ln x_n), \forall x \in \mathcal{X},$$

$$\psi^* (u) = \ln \left( \sum_{i=1}^n e^{u_i} \right), \nabla \psi^* (u) = \text{col} \left( \frac{e^{u_1}}{\sum_{i=1}^n e^{u_i}}, \ldots , \frac{e^{u_n}}{\sum_{i=1}^n e^{u_i}} \right),$$

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

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

$$\left[ \nabla^2 \psi^* \left( \nabla \psi \left( x (t) + \frac{t}{\alpha} \hat{x} (t) \right) \right) \right]_{ij}$$

$$= \left( x_i (t) + \frac{t}{\alpha} \hat{x}_i (t) \right) - \left( x_i (t) + \frac{t}{\alpha} \hat{x}_i (t) \right) \left( x_j (t) + \frac{t}{\alpha} \hat{x}_j (t) \right).$$

Therefore, the APDMD (15) reduces to

$$\begin{cases} \dot{x}_i (t) + \frac{\alpha + 1}{\alpha} \hat{x}_i (t) + (x_i (t) + \frac{t}{\alpha} \hat{x}_i (t)) \left( (\nabla f (x (t)) + \beta A^T (Ax (t) - b) \\
+ A^T \left( \lambda (t) + \frac{t}{\alpha} \hat{\lambda} (t) \right) \right) - (x_i (t) + \frac{t}{\alpha} \hat{x}_i (t))^T (\nabla f (x (t))) \\
+ \beta A^T (Ax (t) - b) + A^T \left( \lambda (t) + \frac{t}{\alpha} \hat{\lambda} (t) \right) \right) \right) = 0, \forall i = 1, \ldots , n, \\
x (t_0) = x_0 \in \mathcal{X}, \hat{x} (t_0) = \hat{x}_0, \lambda (t_0) = \lambda_0, \hat{\lambda} (t_0) = \hat{\lambda}_0, \end{cases}$$

(28)

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

**Case 4:** If $\mathcal{X}$ is a closed and convex set, and its projection operator $P_\mathcal{X}$ has a closed form solution, moreover, setting $\psi (x) = \frac{1}{2} \| x \|^2 + I_X (x)$, one has $\partial_\psi (x) = x + N_\mathcal{X} (x)$ 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:

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

(29)
3.8 APDMD for DCCP (12) in the smooth case

In this subsection, inspired by the ADPDMD (15) 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 APDMD with $\alpha \geq 2$ is given by

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

(30)

Defining $L = L_n \otimes I_m$, where $L_n$ is the Laplacian matrix of graph $G$ and let $x = \text{col}(x_1, \ldots, x_n) \in \mathbb{R}^{nm}$, $\dot{x} = \text{col} (\dot{x}_1, \ldots, \dot{x}_n) \in \mathbb{R}^{nm}$, $\lambda = \text{col}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{nm}$, $\dot{\lambda} = \text{col}(\dot{\lambda}_1, \ldots, \dot{\lambda}_n) \in \mathbb{R}^{nm}$, $\nabla f (x) = \text{col}(\nabla f_1 (x_1), \ldots, \nabla f_n (x_n)) \in \mathbb{R}^{nm}$, $\nabla \psi^* (u_i) = \text{col}(\nabla \psi_i^* (u_1), \ldots, \nabla \psi_i^* (u_n)) \in \mathbb{R}^{nm}$ and $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$ is the Cartesian product of set $\mathcal{X}_i$, $i = 1, \ldots, n$. Then the compact formula of APDMD (30) is given by

\[
\begin{align*}
\dot{x}(t) &= \frac{\sigma}{t} (\nabla \psi^* (u(t)) - x(t)), \\
\dot{u}(t) &= -\frac{\lambda}{\alpha} \left( \nabla f (x(t)) + \beta Lx(t) + Lv(t) \right), \\
\dot{\lambda}(t) &= \frac{\sigma}{\alpha} (v(t) - \lambda(t)), \dot{v}(t) = \frac{\beta}{\alpha} L \nabla \psi^* (u(t)), \\
x(t_0) &= x_0, \nabla \psi^* (u(t_0)) = x_0, \lambda(t_0) = \lambda_0, v(t_0) = v_0.
\end{align*}
\]

(31)

Next, we will illustrate the accelerated convergence rate of the ADPDD (31) by the Lyapunov analysis method.

**Theorem 12.** Suppose Assumption 3.2 holds and let $(x(t), \lambda(t))$ and $(x^*, \lambda^*)$ be a solution trajectory and an optimal solution for ADPDMD (31) and DCCP (12), respectively. Then for any $(x(t_0), u(t_0), \lambda(t_0), v(t_0)) \in \mathcal{X} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$, the following statements are true:

(I): The trajectories $x(t)$ and $\lambda(t)$ of ADPDMD (31) are bounded for any $t \geq t_0 > 0$. 

21
(II): Let \( L_\beta (x(t), \lambda(t)) = f(x(t)) + \frac{\beta}{2} x^T(t) L x(t) + \lambda(t)^T L x(t) \), one has

\[
L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t)) \leq \frac{\alpha^2 V(t_0)}{t^2},
\]

\[
x(t)^T L x(t) \leq \frac{2\alpha^2 V(t_0)}{\beta t^2},
\]

\[
\int_{t_0}^{+\infty} t \left( L_\beta (x(t), \lambda^*) - L_\beta (x^*, \lambda(t)) \right) dt < +\infty,
\]

\[
\int_{t_0}^{+\infty} t \| Ax(t) - b \|^2 dt < +\infty.
\]

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

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

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

**Proof** Consider the following candidate Lyapunov function:

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

with

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

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

3.9 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 \( \sum_{i=1}^{n} A_i x_i = \sum_{i=1}^{n} d_i \in \mathbb{R}^m \) can be equivalently decomposed as \( Ax - d + Ly = 0 \in \mathbb{R}^{nm} \) with \( L = L_n \otimes I_m \) and an auxiliary variable \( y = \text{col} (y_1, ..., y_n) \in \mathbb{R}^{nm} \) by using the properties of the Laplacian matrix of undirected \( G \) (i.e., \( \ker (L_n) = \{ \varsigma 1 | \varsigma \in \mathbb{R} \} \), \( \text{rang} (L_n) = \{ \omega \in \mathbb{R}^n | \omega^T 1 = 0 \} \)).
Therefore, the DEMO (13) is equivalent to:

\[
\min_{x \in \mathbb{R}^{\sum_{i=1}^{n} p_i}, y \in \mathbb{R}^{nm}} f(x) = \sum_{i=1}^{n} f_i(x_i),
\]

\[
\text{s.t. } \bar{A} x - d + Ly = 0, x \in \mathcal{X} = \prod_{i=1}^{n} \mathcal{X}_i,
\]

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

To deal with the modified DEMO (35) with smooth convex objective functions, we propose the following accelerated distributed mirror dynamical (ADMD) approach with \( t \geq t_0 > 0 \) and \( \alpha \geq 2 \) as follows:

\[
\begin{align*}
\dot{x}(t) &= \frac{\alpha}{t} (\nabla \psi^* (u(t)) - x(t)) , \\
\dot{u}(t) &= -\frac{\alpha}{t} \left( \nabla f(x(t)) + \bar{A}^T v(t) \right) , \\
\dot{\lambda}(t) &= \frac{\alpha}{t} (v(t) - \lambda(t)) , \\
\dot{v}(t) &= \frac{\alpha}{t} \left( \bar{A} \psi^* (u(t)) - d - L \lambda(t) + L z(t) \right) , \\
\dot{z}(t) &= -\frac{\alpha}{t} L v(t) .
\end{align*}
\]

\( (36) \)

**Theorem 13.** 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 (36), respectively. Then, for any \( (x(t_0), u(t_0), \lambda(t_0), v(t_0), y(t_0), z(t_0)) \in \mathcal{X} \times \mathbb{R}^{\sum_{i=1}^{n} p_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \), we have the following two statements:

(I): The trajectories \( x(t), \lambda(t) \) and \( y(t) \) of ADMD (36) are bounded for any \( t \geq t_0 > 0. \)

(II): Let \( L_\beta \) with \( \beta = 1, \) i.e., \( L_1 (x(t), \lambda(t), y(t)) \) be \( f(x(t)) - \frac{1}{2} \lambda(t)^T L \lambda(t) \) \( (\bar{A} x(t) - d - L y(t)), \) it follows that

\[
L_1 (x(t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda^*) \leq \frac{\alpha^2 E(t_0)}{t^2}, \quad \lambda(t)^T L \lambda(t) \leq \frac{2 \alpha^2 E(t_0)}{t^2},
\]

\[
\int_{t_0}^{+\infty} \left( L_1 (x(t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda(t)) \right) dt < +\infty.
\]

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

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

with

\[
\begin{align*}
E_1(t) &= \frac{\alpha^2}{2t} \left( L_1 (x(t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda(t)) \right) , \\
E_2(t) &= \frac{\alpha^2}{2t} \left( f(x(t)) - f(x^*) + (\lambda^*)^T (\bar{A} x(t) - d - L y^*) + \frac{\beta}{2} \lambda(t)^T L \lambda(t) \right), \\
E_3(t) &= D_{\psi^*} (u, u^*) = \psi^* (u(t)) - \psi^* (u^*) - \nabla \psi^* (u^*)^T (u(t) - u^*) , \\
E_4(t) &= D_h (z(t), y^*) = \frac{1}{2} \| v(t) - \lambda^* \|^2 ,
\end{align*}
\]

\( (38) \)
Using the same proof steps as (11), we can get

\[
{\dot{E} (t)} = \dot{E}_1 (t) + \dot{E}_2 (t) + \dot{E}_3 (t) + \dot{E}_4 (t) \\
\leq - \frac{(\alpha - 2)t}{\alpha^2} f (x (t)) - f (x^*) + (\lambda^*)^T A (x (t) - x^*) + \frac{1}{2} \lambda^T \mathcal{L} \lambda (t) \\
= - \frac{(\alpha - 2)t}{\alpha^2} (L_1 (x (t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda (t))) \leq 0.
\]

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

(II): Note that \( \dot{E} (t) \leq 0 \), i.e., \( E (t) \) is nonincreasing on \([t_0, +\infty)\). Thus, for any \( t \geq t_0 > 0 \), one has \( \frac{t^2}{\alpha^2} (L_1 (x (t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda (t))) = \dot{E}_1 (t) \leq E (t) \leq E (t_0) \), i.e., \( L_1 (x (t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda (t)) \leq \frac{\alpha^2 E(t_0)}{t^2} \) holds. Furthermore, we have \( \lambda (t)^T \mathcal{L} \lambda (t) \leq \frac{2\alpha^2 E(t_0)}{t^2} \). Thus, the conclusion of (37a) holds.

Since \( \dot{E} (t) \leq - \frac{(\alpha - 2)t}{\alpha^2} (L_1 (x (t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda (t))) \), thus, integrating it from \( t_0 \) to \(+\infty\) and rearranging it, we have

\[
\int_{t_0}^{+\infty} t (L_1 (x (t), y^*, \lambda^*) - L_1 (x^*, y^*, \lambda (t))) dt \\
= \frac{\alpha^2}{\alpha - 2} (E (t_0) - E (+\infty)) \leq \frac{\alpha^2}{\alpha - 2} E (t_0) < +\infty.
\]

Therefore, the conclusion of (37b) holds.

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:

\[
\begin{align*}
\dot{x}^\mu (t) &= \frac{\alpha}{T} \left( \nabla \psi^* (u^\mu (t)) - x^\mu (t) \right), \\
\dot{u}^\mu (t) &= - \frac{T}{\alpha} \nabla_x \bar{f} (x^\mu (t), \mu (t)) + \beta A^T (Ax^\mu (t) - b) + A^T u^\mu (t), \\
\dot{\lambda}^\mu (t) &= \frac{\alpha}{T} (u^\mu (t) - \lambda^\mu (t)), \\
\dot{v}^\mu (t) &= \frac{T}{\alpha} (A \nabla \psi^* (u^\mu (t)) - b (t)), \\
x^\mu (t_0) &= x^\mu_0, u^\mu (t_0) = u^\mu_0, \nabla \psi^* (u^\mu_0) = x^\mu_0, \\
\lambda^\mu (t_0) &= \lambda^\mu_0, v^\mu (t_0) = v^\mu_0, \mu (t_0) = \mu_0,
\end{align*}
\]

where \( t \geq t_0 > 0, \beta > 0, \alpha \geq 2 \) and \( \mu (t) \leq \frac{\mu_0}{1 - \alpha}. \) The superscript \( \mu \) means the trajectories that are obtained by the smoothing dynamical approaches, which is used to distinguish
the trajectories obtained by the dynamical approaches without smoothing approximation, i.e., APDMD (14), ADPDM (31) and ADMD (36). The $\nabla_x \hat{f}(x^\mu(t), \mu(t))$ is the gradient of smoothing function of $\hat{f}(x^\mu(t), \mu(t))$ with respect of $x^\mu(t)$. The SAPDMD (39) is illustrated in (4) (left).

Similar to APDMD (14), the SAPDMD (39) can be equivalent to the following smoothing second-order dynamical approach:

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

$$
\begin{align*}
x^\mu(t_0) = x^\mu_0 \in X, \quad \dot{x}^\mu(t_0) = \dot{x}^\mu_0, \quad \lambda^\mu(t_0) = \lambda^\mu_0, \quad \dot{\lambda}^\mu(t_0) = \dot{\lambda}^\mu_0.
\end{align*}
$$

Figure 4: (left) The demonstration of SAPDMD (39). (right) The demonstration of the proof of feasibility of SAPDMD (39).

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, Cabot and Paoli (2007); Apidopoulos et al. (2018); He et al. (2017), Moreau-Yosida regularization method Balavoine et al. (2013); Attouch et al. (2021) and directional derivative method Su et al. (2016); Fazlyab et al. (2017).

Compared with the dynamical approaches based on techniques mentioned above for solving nonsmooth optimization problems, our proposed SAPDMD (39) based on smoothing approximation in (2.6) has the following differences and preponderances.

- The solutions may be different. The SAPDMD (39) has a strong global solution, which is similar to that in (9). In Apidopoulos 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 (39) 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); 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 $\varepsilon$-solution, i.e., $f(x^\varepsilon) - \min f(x) \leq \varepsilon$, where $\varepsilon$ is a small constant, but it’s not equal to 0.

- The proposed SAPDMD (39) does not require solving some subproblems. However, the accelerated dynamic approaches in Balavoine et al. (2013); Attouch et al. (2021) need to use the Moreau-Yosida approximation $f_\rho(x) = \min_{x \in X} \{ f(\xi) + \frac{1}{2\rho} \| x - \xi \|^2 \}$; the accelerated dynamic approaches Su et al. (2016); Fazlyab et al. (2017) need to utilize $d(x; \dot{x}) = \arg \max_{f \in \partial f} f^T \dot{x}$, 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 (39) can be easily guaranteed by the Cauchy-Lipschitz-Picard theorem. However, the uniqueness of the solutions in directional derivative method He et al. (2017) and differential inclusion method Su et al. (2016); Fazlyab 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 (39) by the same way as in (3.3).

**Lemma 14.** For any $(x^\mu(t_0), u^\mu(t_0), \lambda^\mu(t_0), \nu^\mu(t_0)) \in X \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, then the variable $x^\mu(t)$ of SAPDMD is always in $X$, $\forall t \geq t_0 > 0$, i.e., the feasibility of $x(t)$ is satisfied.

**Proof** The proof follows the same arguments as in (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 (1) with nonsmooth objective functions and we will once again take into account solutions (similar to (9)) to this problem.

**Theorem 15.** For any initial values $(x^\mu(t_0), u^\mu(t_0), \lambda^\mu(t_0), \nu^\mu(t_0)) \in X \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, there exists a unique strong global solution of SAPDMD (39).

**Proof** Let $Y^\mu(t) = (x^\mu(t), u^\mu(t), \lambda^\mu(t), \nu^\mu(t))$, then the SAPDMD (39) is equivalent to

\[
\begin{align*}
\dot{Y}^\mu(t) &= H(t, Y^\mu(t)) \\
Y^\mu(t_0) &= (x^\mu(t_0), u^\mu(t_0), \lambda^\mu(t_0), \nu^\mu(t_0)),
\end{align*}
\]
where

\[
H (t, Y^\mu) = \left( \frac{\alpha}{t} (\nabla \psi^* (u^\mu (t)) - x^\mu (t)) \right), \quad -\frac{t}{\alpha} \left( \nabla_x \tilde{f} (x^\mu (t), \mu (t)) - A^T \psi^\mu (t) \right), \quad \frac{\alpha}{t} (\psi^\mu (t) - \lambda^\mu (t)), \quad \frac{t}{\alpha} (A \nabla \psi^* (u^\mu (t)) - b) \right) .
\]  

(42)

To prove the existence and uniqueness of the strong global solution \( Y^\mu (t) \) generated by SAPDMD (39) by the Cauchy-Lipschitz-Picard theorem Bolte (2003), the following conditions need to be satisfied:

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

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

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

\[
\left\| F (t, Y^\mu (t)) - F (t, \hat{Y}^\mu (t)) \right\| \leq \left( \frac{t^2}{\alpha^2} \left( 4 \beta^2 \delta_{\max} (A^T A)^2 + \frac{4t^2}{\mu^2 (t)} \right) + \frac{\alpha^2 \left( 6 + 2l^2 \psi^* \right)}{t^2} \right) \left\| Y^\mu (t) - \hat{Y}^\mu (t) \right\| .
\]

By using the notation \( l (t) = \left( \frac{t^2 \left( 4 \beta^2 \delta_{\max} (A^T A)^2 + \frac{4t^2}{\mu^2 (t)} + \delta_{\max} (A^T A) l_{\psi^*} + 2 \delta_{\max} (A^T A) \right)}{\alpha^2} + \frac{\alpha^2 \left( 6 + 2l^2 \psi^* \right)}{t^2} \right)^{\frac{1}{2}} \), one has

\[
\left\| H (t, Y^\mu (t)) - H (t, \hat{Y}^\mu (t)) \right\| \leq l (t) \left\| Y^\mu (t) - \hat{Y}^\mu (t) \right\| .
\]

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

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

\[
\int_{t_0}^{T} \left\| H (t, Y^\mu (t)) \right\| dt \leq \sqrt{\left\| \nabla \psi^* (u (t)) \right\|^2 + \left\| x (t) \right\|^2 + \left\| \lambda (t) \right\|^2 + \left\| v (t) \right\|^2 + \left\| x^* \right\|^2 + \left\| \nabla_x \tilde{f} (x^\mu (t), \mu (t)) \right\|^2} \times \left( \frac{8 \alpha^2}{t^2} + \frac{2t^2}{\alpha^2} \left( 4 \delta_{\max} (A^T A) + 4 \beta^2 \delta_{\max} (A^T A)^2 + 2 \right) \right),
\]

and the conclusion holds by employing the continuity of the function

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

The existence and uniqueness of \( Y^\mu (t) \) to the dynamical system (42) can be guaranteed by the Cauchy-Lipschitz-Picard theorem, consequently, the existence and uniqueness of the trajectories of SAPDMD (39) also hold.
4.3 The accelerated convergence of the SAPDMD

In this subsection, we will illustrate the accelerated convergence properties of the SAPDMD (39) 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 16.** Suppose that Assumption 3.1 holds and the objective function is nonsmooth convex. Let $(x(t), y(t))$ and $(x^*, y^*)$ be a solution trajectory and an optimal solution of SAPDMD (39), respectively. Defining a Lyapunov function as $\hat{L}_\beta(x^\mu(t), \lambda^\mu(t)) = \hat{f}(x^\mu(t), \mu(t)) + \lambda^\mu(t)^T(Ax^\mu(t) - b) + \frac{\beta}{2} \|Ax^\mu(t) - b\|^2$ with $\beta \geq 0$, then, for any $(x^\mu(t), \lambda^\mu(t)) \in X \times \mathbb{R}^m$, we have

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

(II): Then, we have

\[
0 \leq \dot{\hat{L}}_\beta(x^\mu(t), \lambda^*) - \hat{L}_\beta(x^*, \lambda^\mu(t)) + 4\kappa \mu(t) \leq \frac{\alpha^2 V(t_0)}{t^2}, \tag{43a}
\]

\[
\|Ax^\mu(t) - b\|^2 \leq \frac{2\alpha^2 V(t_0)}{\beta t^2}, \tag{43b}
\]

\[
\int_{t_0}^{+\infty} \|Ax^\mu(t) - b\|^2 dt \leq +\infty, \tag{43c}
\]

\[
\int_{0}^{+\infty} t \left(\dot{\hat{L}}_\beta(x^\mu(t), \lambda^*) - \hat{L}_\beta(x^*, \lambda^\mu(t)) + 2\kappa \mu(t)\right) dt < +\infty, \tag{43d}
\]

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

\[
0 \leq f(x^\mu(t)) - f(x^*) + \|Ax^\mu(t) - b\| \leq \frac{\alpha^2 V(t_0)}{t^2}; \tag{44a}
\]

\[
- \frac{\alpha \sqrt{2V(t_0)}}{t \sqrt{\beta}} \leq f(x^\mu(t)) - f(x^*) \leq \frac{\alpha^2 V(t_0)}{t^2}. \tag{44b}
\]

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

\[
V(t) = V_1(t) + V_2(t) + V_3(t), \tag{45}
\]
with

\[
\begin{align*}
V_1(t) &= \frac{\mu}{\alpha^2} \left( \hat{\mathcal{L}}_\beta(x^t(t), \lambda^t) - \hat{\mathcal{L}}_\beta(x^*, \lambda^t(t)) + 4\kappa \mu(t) \right) \\
&= \frac{\mu}{\alpha^2} \left( f(x^t(t), \mu(t)) - f(x^*, \mu(t)) + 4\kappa \mu(t) \\
&\quad + \frac{\beta}{2} ||Ax^t(t) - b||^2 + (\lambda^*)^T A(x^t(t) - x^*) \right) \\
V_2(t) &= D_{\psi^*}(u^t(t), u^*) = \psi^*(u^t(t)) - \psi^*(u^*) - \nabla \psi^*(u^*)^T (u^t(t) - u^*) \\
V_3(t) &= D_h(v^t(t), \lambda^t) = \frac{1}{2} ||v^t(t) - \lambda^t||^2.
\end{align*}
\]

Note that the function \(V_1(t)\) is positive for any \(t \geq t_0 > 0\) since

\[
\begin{align*}
\hat{\mathcal{L}}_\beta(x^t(t), \lambda^*) - \hat{\mathcal{L}}_\beta(x^*, \lambda^t(t)) + 4\kappa \mu(t) \\
\geq f(x^t(t)) - f(x^*) + (\lambda^*)^T A(x^t(t) - x^*) + \frac{\beta}{2} ||Ax^t(t) - b||^2 + 2\kappa \mu(t) \\
= \mathcal{L}_\beta(x^t(t), \lambda^*) - \mathcal{L}_\beta(x^*, \lambda^t(t)) + 2\kappa \mu(t) \geq 0,
\end{align*}
\]

where the first equality holds from \(\hat{f}(x^t(t), \mu(t)) - f(x^t(t)) + \kappa \mu(t) \geq 0\), \(f(x^*) - f(x^*, \mu(t)) + \kappa \mu(t) \geq 0\) in (6) \(\hat{u}\), and the second inequality is satisfied due to (9) in (2.2). In addition, since \(\psi^* (\cdot)\) and \(h(\cdot) = \frac{1}{2} \|\cdot\|^2\) are convex, such that \(D_{\psi^*}(u^t(t), u^*) \geq 0\), \(D_h(v^t(t), \lambda^t) \geq 0\). Thus, we end up with \(V(t) \geq 0\).

The derivatives of \(V_1(t), V_2(t)\) and \(V_3(t)\) along the trajectory of SAPDMD (39) satisfy

\[
\begin{align*}
\dot{V}_1(t) &= \frac{2t}{\alpha^2} \left( \dot{f}(x^t(t), \mu(t)) - f(x^*, \mu(t)) + (\lambda^*)^T A(x^t(t) - x^*) \\
&\quad + 4\kappa \mu(t) + \frac{\beta}{2} ||Ax^t(t) - b||^2 \right) + \frac{t}{\alpha} \left( \nabla \psi^*(u^t(t)) - \nabla \psi^*(u^*) \right)^T \\
&\quad \times \left( \nabla_x \dot{f}(x^t(t), \mu(t)) + \beta A^T (Ax^t(t) - b) + A^T \lambda^* \right) \\
&\quad + \frac{t}{\alpha} \left( \nabla_x \dot{f}(x^t(t), \mu(t)) + \beta A^T (Ax^t(t) - b) + A^T \lambda^* \right)^T \\
&\quad \times \left( \nabla \psi^*(u^*) - x^t(t) \right) + \frac{t^2}{\alpha^2} \kappa \dot{\mu}(t) \\
&\quad + \frac{t^2}{\alpha^2} \left( \nabla_{\mu} \dot{f}(x^t(t), \mu(t)) + \nabla_{\mu} \dot{f}(x^*, \mu(t)) + 2\kappa \dot{\mu}(t) \right),
\end{align*}
\]

\[
\begin{align*}
\dot{V}_2(t) &= - \frac{t}{\alpha} \left( \nabla \psi^*(u^t(t)) - \nabla \psi^*(u^*) \right)^T (A^T u^t(t) - A^T \lambda^*) \\
&\quad - \left( \nabla_x \dot{f}(x^t(t), \mu(t)) + \beta A^T A(x^t(t) - x^*) + A^T \lambda^* \right)^T \\
&\quad \times \frac{t}{\alpha} \left( \nabla \psi^*(u^t(t)) - \nabla \psi^*(u^*) \right),
\end{align*}
\]

\[
\dot{V}_3(t) = (v^t(t) - \lambda^*)^T \dot{v}^t(t) = \frac{t}{\alpha} \left( v^t(t) - \lambda^* \right)^T A (\nabla \psi^*(u^t(t)) - x^*).
\]
Combining (46), (47), (48) and rearranging it yields
\[
\dot{V}(t) = \dot{V}_1(t) + \dot{V}_2(t) + \dot{V}_3(t)
\]
\[
\leq \frac{2t}{\alpha^2} \left( \hat{f}(x^\mu(t), \mu(t)) - \hat{f}(\hat{x}^\mu, \mu(t)) + \frac{\beta}{2} \|A x^\mu(t) - b\|^2 + 4\kappa_f \mu(t) 
\right.
\]
\[
+ \langle \lambda^\mu \rangle^T A (x^\mu(t) - x^*) \right) - \frac{t}{\alpha} \left( \hat{f}(x^\mu(t), \mu(t)) - \hat{f}(\hat{x}^\mu, \mu(t)) 
\right.
\]
\[
+ \beta \|A x^\mu(t) - b\|^2 + \langle \lambda^\mu \rangle^T A (x^\mu(t) - x^*) \right) + \frac{2t^2}{\alpha^2} \kappa_f \hat{\mu}(t) 
\]
\[
\leq \frac{(2 - \alpha) t}{\alpha^2} \left( \hat{f}(x^\mu(t), \mu(t)) - \hat{f}(\hat{x}^\mu, \mu(t)) + \frac{\beta}{2} \|A x^\mu(t) - b\|^2 
\right.
\]
\[
+ \langle \lambda^\mu \rangle^T A (x^\mu(t) - x^*) + \frac{\beta}{2} \|A x^\mu(t) - b\| 
\right)
\]
\[
- \frac{t^2}{2\alpha} \|A x^\mu(t) - b\|^2 
\leq 0, \forall x^\mu(t) \in \mathcal{X}, t \geq t_0 > 0,
\]
where the first inequality holds since $\nabla_\mu \hat{f}(x^\mu(t), \mu(t)) + \nabla_\mu \hat{f}(x^\mu(t)) + 2\kappa_f \geq 0$ and $\hat{\mu}(t) \leq 0$, the second inequality is satisfied because of the convexity of $\hat{f}(x^\mu(t), \mu(t))$ of $x^\mu(t) \in \mathcal{X}$ with any fixed $\mu(t) \in [0, \bar{\mu}]$, the third inequality holds from $\mu(t) \leq \frac{\mu_0}{2\alpha}$ (i.e., $\hat{\mu}(t) \leq -\mu_0 (2\alpha) t^{(-2\alpha - 1)}$ and $\frac{\mu_0}{2\alpha} \kappa_f \geq 0$ imply $\frac{\mu_0}{2\alpha} \kappa_f \hat{\mu}(t) \leq \frac{\mu_0}{2\alpha} \kappa_f \mu_0 t^{-2\alpha} - \frac{\mu_0}{2\alpha} \kappa_f \mu_0 (2\alpha) t^{-2\alpha - 1} = \frac{\mu_0}{2\alpha} \kappa_f \mu_0 t^{-2\alpha - 2\alpha} = 0$), and the last inequality holds since $\alpha \geq 2$ and $\hat{L}_\beta (x^\mu(t), \lambda^\mu) - \hat{\hat{L}}_\beta (x^\mu, \lambda^\mu(t)) + 4\kappa_f \mu(t) \geq 0$.

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

(II): From (49), we have that $\dot{V}(t) \leq 0$, i.e., $V(t)$ is nonincreasing on $t \in [t_0, +\infty)$. It yields to
\[
\frac{t^2}{2\alpha} \left( \hat{L}_\beta (x^\mu(t), \lambda^\mu) - \hat{\hat{L}}_\beta (x^\mu, \lambda^\mu(t)) + 4\kappa_f \mu(t) \right) = V_1(t) \leq V(t) \leq V(t_0), \text{ i.e., (43a) holds.}
\]
Furthermore, the (43a) implies $\|A x^\mu(t) - b\|^2 \leq \frac{2\alpha^2 V(t_0)}{\beta^2}$, i.e., (43b) is satisfied.

Note that $\dot{V}(t) \leq -\frac{(2 - \alpha) t}{\alpha^2} \left( \hat{L}_\beta (x^\mu(t), \lambda^\mu) - \hat{\hat{L}}_\beta (x^\mu, \lambda^\mu(t)) + 4\kappa_f \mu(t) \right) - \frac{t^2}{2\alpha} \|A x^\mu(t) - b\|^2$ in (49), and integrating it from $t_0$ to $+\infty$ and rearranging it, we have
\[
\frac{\alpha - 2}{\alpha^2} \int_{t_0}^{+\infty} t \left( \hat{L}_\beta (x^\mu(t), \lambda^\mu) - \hat{\hat{L}}_\beta (x^\mu, \lambda^\mu(t)) + 4\kappa_f \mu(t) \right) dt
\]
\[
= V(t_0) - V(+\infty) \leq V(t_0) < +\infty,
\]
\[
\frac{\beta}{2\alpha} \int_{t_0}^{+\infty} t \|A x^\mu(t) - b\|^2 dt = V(t_0) - V(+\infty) \leq V(t_0) < +\infty,
\]
which implies both (43c) and (43d) hold.
(III): The equations \( \lambda^* = \begin{cases} 0, & \text{if } Ax^\mu(t) - b = 0, \\ \frac{\alpha \sqrt{2V(t_0)}}{\sqrt{\beta t}} \left( \frac{Ax^\mu(t) - b}{\|Ax^\mu(t) - b\|} \right), & \text{if } Ax^\mu(t) - b \neq 0, \end{cases} \) and (43a) give \( \|Ax^\mu(t) - b\| \leq \alpha \sqrt{2V(t_0)} / \sqrt{\beta t} \).

According to the upper-bound of (9) in (2.2) with \( \beta = 0 \), one has
\[
\mathcal{L}_0(x^\mu(t), \lambda^*) - \mathcal{L}_0(x^*, \lambda^*(t)) = f(x^\mu(t)) - f(x^*) \geq -(Ax^\mu(t) - b)^T \lambda^*, \forall (x^\mu(t), \lambda^*(t)) \in \mathcal{X} \times \mathbb{R}^m,
\]
and from (iv) in (6), one also has
\[
\hat{f}(x^\mu(t), \mu(t)) - \hat{f}(x^*, \mu(t)) + 4\kappa f \mu(t)
\geq f(x^\mu(t)) - f(x^*) \geq -(Ax^\mu(t) - b)^T \lambda^*
\geq -\|Ax^\mu(t) - b\| \geq -\frac{\alpha \sqrt{2V(t_0)}}{t \sqrt{\beta}}, \forall (x^\mu(t), \lambda^*(t)) \in \mathcal{X} \times \mathbb{R}^m.
\]

Since \( \|Ax^\mu(t) - b\| \geq 0 \) and (43a) imply \( f(x^\mu(t)) - f(x^*) \leq \frac{\alpha^2 V(t_0)}{t^2} \), further, we deduce for any \( t \geq t_0 > 0 \) that \( -\frac{\alpha \sqrt{2V(t_0)}}{t \sqrt{\beta}} \leq f(x^\mu(t)) - f(x^*) \leq \frac{\alpha^2 V(t_0)}{t^2} \) is true. Therefore, we can obtain the conclusion (III).

Similar to the smooth case, the SAPDMD (39) can also be used to address the DCCP (12) and DEMO (13) with nonsmooth convex objectives.

### 4.4 SAPDMD for DCCP in the nonsmooth case

Based on the SAPDMD (39), a smoothing accelerated distributed primal-dual mirror dynamical (SAPDMD) approach for DCCP (12) with nonsmooth convex objective functions is given by:

\[
\begin{align*}
\dot{x}^\mu(t) & = \frac{\alpha}{t} \left( \nabla \hat{\psi}^*(u^\mu(t)) - x^\mu(t) \right), \\
\dot{u}^\mu(t) & = -\frac{1}{\alpha} \left( \nabla_x \hat{f}(x^\mu(t), \mu(t)) + \beta Lx^\mu(t) + L(u^\mu(t)) \right), \\
\dot{\lambda}^\mu(t) & = \frac{\alpha}{\beta} \left( v^\mu(t) - \lambda^\mu(t) \right), \quad \dot{v}^\mu(t) = \frac{\lambda}{\alpha} L\nabla \psi^*(u^\mu(t)), \\
x^\mu(t_0) & = x^\mu_0, \quad \nabla \psi^*(u^\mu(t_0)) = x^\mu_0, \quad \lambda^\mu(t_0) = \lambda^\mu_0, \quad v^\mu(t_0) = v^\mu_0, \quad \mu_0 = \bar{\mu} > 0,
\end{align*}
\]

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

### 4.5 The accelerated convergence of the SAPDMD

Now, let’s discuss accelerated convergence properties of SAPDMD (50) with the help of Lyapunov analysis tool.

**Theorem 17.** Suppose that Assumption 3.2 holds and the objective function is nonsmooth. Let \((x^\mu(t), \lambda^\mu(t))\) and \((x^*, \lambda^*)\) be a solution trajectory and an optimal solution of SAPDMD (50), respectively. Let \( \mathcal{L}_\beta(x^\mu(t), \lambda^\mu(t)) = \hat{f}(x^\mu(t), \mu(t)) + \frac{\alpha}{2}(x^\mu(t))^T L(x^\mu(t)) + (\lambda^\mu(t))^T Lx^\mu(t) \) with \( \beta \geq 0 \), then, for \( \alpha \geq 2 \), one has
(I): $x^\mu(t)$, $\lambda^\mu(t)$, $u^\mu(t)$, $v^\mu(t)$ are bounded for any $t \geq t_0 > 0$.

(II): Then, we have

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

(51a)

$$x^\mu(t) L x^\mu(t) \leq \frac{2\alpha^2 \mathcal{V}(t_0)}{\beta t^2},$$

(51b)

$$\int_{t_0}^{+\infty} x^\mu(t) L x^\mu(t) \leq +\infty,$$

(51c)

$$\int_{t_0}^{+\infty} t \left( \hat{L}_\beta (x^\mu(t), \lambda^*) - \hat{L}_\beta (x^*, \lambda^\mu(t)) + 2\kappa \dot{\mu}(t) \right) dt < +\infty,$$

(51d)

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

$$f(x^\mu(t)) - f(x^*) + ||L x^\mu(t)|| \leq \frac{\alpha^2 \mathcal{V}(t_0)}{t^2};$$

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

(52a, 52b)

Proof Construct the candidate smoothing Lyapunov function as

$$\mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2 + \mathcal{V}_3,$$

(53)

with

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

Following similar steps as the proof in (16), the conclusions can be obtained easily. Due to space limitations, we omit the proof here. 

\[\square\]

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 (39) and distributed consensus theorem is given by

\[
\begin{align*}
\dot{x}^\mu(t) &= \frac{\alpha}{\beta} (\nabla \psi^* (u^\mu(t)) - x^\mu(t)), \\
\dot{u}^\mu(t) &= -\frac{\alpha}{\beta} \left( \nabla_2 f(x^\mu(t), \mu(t)) + \nabla \psi^* (u^\mu(t)) \right), \\
\dot{\lambda}^\mu(t) &= \frac{\alpha}{\beta} (v^\mu(t) - \lambda^\mu(t)), \\
\dot{v}^\mu(t) &= \frac{\alpha}{\beta} (\dot{\lambda}^\mu(t) - y^\mu(t), \\
\dot{z}^\mu(t) &= -\frac{\alpha}{\beta} L \dot{x}^\mu(t), \\
x^\mu(t_0) &= x^\mu_0, u^\mu(t_0) = u^\mu_0 \text{ with } \nabla \psi^* (u_0) = x^\mu_0 \in \mathcal{X}, \\
\lambda^\mu(t_0) &= \lambda^\mu_0, v^\mu(t_0) = v^\mu_0, y^\mu(t_0) = y^\mu_0, z^\mu(t_0) = z^\mu_0, \mu_0 = \bar{\mu} > 0,
\end{align*}
\]

(54)
where \( t \geq t_0 > 0, \beta > 0, \alpha \geq 2, \) and \( \mu (t) \leq \frac{\mu_0}{t^\beta}. \)

The accelerated convergence properties of SADMD (54) will be demonstrated in the following theorem.

**Theorem 18.** Suppose that Assumption 3.2 holds, except that the objective function is non-smooth. Let \((x^\mu (t), \lambda^\mu (t), y^\mu (t))\) and \((x^*, \lambda^*, y^*)\) be a solution trajectory and an optimal solution of SADMD (54), respectively. Then, for any initial values \((x^\mu (t_0), u^\mu (t_0), \lambda^\mu (t_0), \nu^\mu (t_0), y^\mu (t_0), z^\mu (t_0)) \in \mathcal{X} \times \mathbb{R}^{\sum_{i=1}^n} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}, \) the following statements are true.

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

(II): Let

\[ \hat{L}_1 (x, \lambda, y) = \check{f} (x^\mu (t), \mu (t)) - \frac{1}{2} \lambda^\mu (t)^T L \lambda^\mu (t) + \lambda^\mu (t)^T (\tilde{A} x^\mu (t) - d - L y^\mu (t)), \]

then, it follows

\[
\begin{align}
\hat{L}_1 (x^\mu (t), y^*, \lambda^*) - \check{L}_1 (x^\mu (t), y^*, \lambda^\mu (t)) + 4 \kappa_j \mu (t) & \leq \frac{\alpha^2 E (t_0)}{t^2}, \\
\lambda^\mu (t)^T L \lambda^\mu (t) & \leq \frac{2 \alpha^2 E (t_0)}{t^2}, \\
\int_{t_0}^{+\infty} t \left( \hat{L}_1 (x^\mu (t), y^*, \lambda^*) - \check{L}_1 (x^\mu (t), y^*, \lambda^\mu (t)) + 4 \kappa_j \mu (t) \right) dt & < +\infty.
\end{align}
\]

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

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

with

\[
\begin{align}
E_1 (t) &= \frac{t^2}{\alpha^2} \left( \hat{L}_1 (x^\mu (t), y^*, \lambda^*) - \check{L}_1 (x^\mu (t), y^*, \lambda^\mu (t)) + 4 \kappa_j \mu (t) \right), \\
E_2 (t) &= \frac{t^2}{\alpha^2} \left( \hat{f} (x^\mu (t), \mu (t)) - \check{f} (x^\mu (t), \mu (t)) + \frac{1}{2} \lambda^\mu (t)^T L \lambda^\mu (t), \\
&\quad + 4 \kappa_j \mu (t) + (\lambda^*)^T (\tilde{A} x^\mu (t) - d - L y^*) \right), \\
E_3 (t) &= D_{\psi^*} (u^\mu (t), u^*) - \psi^* (u^\mu (t)) - \psi^* (u^*) - \nabla \psi^* (u^*)^T (u^\mu (t) - u^*), \\
E_4 (t) &= D_h (z^\mu (t), y^*) = \frac{1}{2} \| z^\mu (t) - y^* \|^2,
\end{align}
\]

where \( D_{\psi^*} (u, u^*) \) and \( D_h (v, \lambda^*) \) are the Bregman divergences associated with \( \psi^* \) for \( u, u^* \), \( h \) for \( v, \lambda^* \) and \( z, y^* \).

The time derivative of \( E (t) \) with \( \nabla \psi^* (u^*) = x^* \) is

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

\[
\leq \frac{2 t}{\alpha^2} \left( \hat{f} (x^\mu (t), \mu (t)) - \check{f} (x^\mu (t), \mu (t)) + (\lambda^*)^T \tilde{A} (x^\mu (t) - x^*) + 4 \kappa_j \mu (t) \right)
\]

\[
+ \frac{1}{2} \lambda^\mu (t)^T L \lambda^\mu (t) - \frac{t}{\alpha} (L \lambda^\mu (t))^T \lambda^\mu (t) + \frac{4 t}{\alpha} \kappa_j \mu (t) + \frac{2 t^2}{\alpha^2} \kappa_j \mu (t)
\]

33
\[-\frac{t}{\alpha} \left( \hat{f} (x^\mu (t), \mu (t)) - \hat{f} (x^*, \mu (t)) + (\lambda^*)^T \bar{A} (x^\mu (t) - x^*) + 4 \kappa \dot{j} \mu (t) \right) \]
\[
\leq -\frac{(\alpha - 2) t}{\alpha^2} \left( \hat{f} (x^\mu (t), \mu (t)) - \hat{f} (x^*, \mu (t)) + (\lambda^*)^T \bar{A} (x^\mu (t) - x^*) + 4 \kappa \dot{j} \mu (t) \right)
\]
\[
+ 4 \kappa \dot{j} \mu (t) + \frac{1}{2} \lambda^T (t) \bar{L} \lambda (t) \]
\[
= -\frac{(\alpha - 2) t}{\alpha^2} \left( \bar{L}_1 (x^\mu (t), y^*, \lambda^*) - \bar{L}_1 (x^*, y^*, \lambda^* (t)) + 4 \kappa \dot{j} \mu (t) \right) \leq 0,
\]
where the first inequality holds because of \( \nabla \mu \hat{f} (x^\mu (t), \mu (t)) \) with any fixed \( \mu (t) \in [0, \bar{\mu}] \) and the property of \( \mu (t) \leq \frac{\mu_0}{t \alpha^2} \) (i.e., \( \dot{\mu} (t) \leq -\mu_0 (2 \alpha) t^{-2 \alpha - 1} \)) and \( \frac{t^2 \sigma^2 \kappa \dot{j}}{\alpha} \geq 0 \) implies \( \frac{2t}{\alpha} \kappa \dot{j} \mu (t) + \frac{t^2 \sigma^2 \kappa \dot{j} \mu (t)}{\alpha} \leq \frac{2t}{\alpha} \kappa \dot{j} \mu_0 t^{-2 \alpha} - \frac{t^2 \sigma^2 \kappa \dot{j} \mu_0 (2 \alpha) t^{-2 \alpha - 1}}{\alpha} = \frac{2t}{\alpha} \kappa \dot{j} \mu_0 t^{-2 \alpha} - \frac{2t^2 \sigma^2 \kappa \dot{j} \mu_0 t^{-2 \alpha}}{\alpha^2} = 0 \), and the last inequality is established since \( \alpha \geq 2 \) and \( \bar{L}_1 (x^\mu (t), \lambda^*) - \bar{L}_1 (x^*, \lambda^* (t)) + 4 \kappa \dot{j} \mu (t) \geq 0 \).

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

(II): From (57), we have that \( \dot{E} (t) \leq 0 \), i.e., \( E (t) \) is nonincreasing on \([t_0, +\infty)\). Thus, for any \( t \geq t_0 > 0 \), one has \( \bar{L}_1 (x^\mu (t), y^*, \lambda^*) - \bar{L}_1 (x^*, y^*, \lambda^* (t)) + 4 \kappa \dot{j} \mu (t) = \bar{E}_1 (t) \leq E (t) \leq E (t_0) \), i.e., \( \bar{L}_1 (x (t), y^*, \lambda^*) - \bar{L}_1 (x^*, y^*, \lambda (t)) + 4 \kappa \dot{j} \mu (t) \leq \frac{\alpha^2 E (t_0)}{t^2} \) (55a) holds. In addition, the (55a) implies \( \lambda^\mu (t)^T \bar{L} \lambda^\mu (t) \leq \frac{2t^2 E (t_0)}{t^2} \), i.e., the (55b) holds.

Since \( \dot{E} (t) \leq -\frac{(\alpha - 2) t}{\alpha^2} \left( \bar{L}_1 (x^\mu (t), y^*, \lambda^*) - \bar{L}_1 (x^*, y^*, \lambda^* (t)) + 4 \kappa \dot{j} \mu (t) \right) \), thus, integrating it from \( t_0 \) to \( +\infty \) and rearranging it, we have
\[
\int_{t_0}^{+\infty} t \left( \bar{L}_1 (x^\mu (t), y^*, \lambda^*) - \bar{L}_1 (x^*, y^*, \lambda^* (t)) + 4 \kappa \dot{j} \mu (t) \right) dt
\]
\[
= \frac{\alpha^2}{\alpha - 2} (E (t_0) - E (+\infty)) \leq \frac{\alpha^2}{\alpha - 2} E (t_0) < +\infty,
\]
therefore, the proof of (55c) is completed.

5. Numerical experiment

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

**Example 2. Logistic regression** Attouch et al. (2021): Consider the problem (1) as follows:

\[
\min f(x) = \log \left( 1 + \exp \left( - (1,1,1)^T x \right) \right)
\]

\[
s.t. \ Ax = b, x \in \mathcal{X},
\]

where \( A = \begin{bmatrix} 0.2,1,1,2 \\ 0,1,0.5,1 \end{bmatrix} \), \( b = \operatorname{col}(1,1) \), \( \mathcal{X} = \left\{ x \in \mathbb{R}^4 \mid \sum_{i=1}^{4} x_i = 1 \right\} \). The objective function is convex (but not strongly convex) and smooth, and it is a very popular regularization in machine learning. Applying APDMD (39) with Kullback-Leibler divergence to address problem (58). (5) (left) shows the trajectories of estimates for \( x(t) \) of APDMD (39) versus time; (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.

**Figure 5:** APDMD (14) with \( \alpha = 2,4,6 \). (left) Trajectories of \( x(t) \) for problem (58). (middle) Error of \( |f(x(t)) - f(x^*)| \) under APDMD (39) for problem (58). (right) Error of \( \|Ax(t) - b\| \) under APDMD (39) for problem (58).

**Example 3. Distributed Logistic regression:** Consider a distributed problem (12) as follows:

\[
\min f(x) = \sum_{i=1}^{n} \log \left( 1 + \exp \left( - \left( i - 1, \frac{i}{2}, i, i + 1 \right)^T x_i \right) \right)
\]

\[
s.t. Lx = 0, x \in \prod_{i=1}^{n} \mathcal{X}_i,
\]

where \( x = (x_1, ..., x_n)^T \in \mathbb{R}^{4n} \). Setting \( n = 4 \) and \( \mathcal{X}_1 = \{ x_1 \in \mathbb{R}^4 \mid 1^T x = 1 \} \), \( \mathcal{X}_2 = \{ x_2 \in \mathbb{R}^4 \mid 1^T x = 1 \} \), \( \mathcal{X}_3 = \{ x_3 \in \mathbb{R}^4 \mid \|x_3 - (0.1, 0.2, 0.5, 0.8)^T\| \leq 2 \} \) and \( \mathcal{X}_4 = \{ x_4 \in \mathbb{R}^4 \mid (1, 1, 1, 1)x_4 \leq 4 \} \). Note that the objective function in (59) is convex (but not strongly convex)
and smooth, which satisfies the requirement of problem (12) in the smooth case. Applying ADPDMD (31) with 4 agents (connected as a ring) to solve problem (59), i.e., for agent 1, using Kullback-Leibler divergence; for agent 2, applying Itakura-Saito divergence; for agent 3, adopting projection operator of Sphere set; for agent 4, utilizing projection operator of half-space set. (6) (left) shows the trajectories of $x(t)$ under ADPDMD (31) for problem (59) are uniformly convergent and globally asymptotically stable, i.e., $x_1 = x_2 = ... = x_5$; (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 $\sqrt{x(t)^T Lx(t)}$ all converge at the predicted rates.

Figure 6: ADPDMD (31) with $\alpha = 3$. (left) Trajectories of $x(t)$ for problem (59). (middle) Error of $|f(x(t)) - f(x^*)|$ under ADPDMD (31) for problem (59). (right) Error of $\sqrt{x(t)^T Lx(t)}$ under ADPDMD (31) for problem (59).

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

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

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

(60)

where $\mathcal{X}_i = \{x_i \in \mathbb{R}^m | i + 2 \leq x_{i,j} \preceq i + 3, j = 1, ..., m\}, i = 1, ..., n$, $b = \text{col}(7, ..., 7) \in \mathbb{R}^{nm}$ and $y \in \mathbb{R}^{nm}$ is an auxiliary variable. Let $n = 10$, $m = 5$ and $A_i \in \mathbb{R}^{10 \times 10}$ be a positive semi-definite matrix generated by standard Gaussian distribution. Let every part $x_i$ be an agent, and 10 agents are connected as a ring. Applying the ADMD (36) to solve problem (60) and the experimental results are shown in (7). In (7) (left), the dual variable $\lambda$ is uniformly convergent and globally asymptotically stable, i.e., $\lambda_1 = \lambda_2 = ... = \lambda_{10}$; (7) (middle and right) illustrate $|f(x^\mu(t)) - f(x^*)|$ and $\sqrt{\lambda(t)^T L\lambda(t)}$ of ADMD (36) are convergent at the predicted rates.
5.2 In the nonsmooth case

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

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

where \( \mathcal{X} = \{ \mathbf{x} \in \mathbb{R}^{40} | x_i \geq 0, i = 1, ..., 40 \} \) and \( \mathbf{A} \in \mathbb{R}^{10 \times 40} \) is an orthogonal Gaussian matrix. The objective function in (61) is convex (but not strongly convex) and nonsmooth as required. Applying the SAPDMD (39) and let \( \psi(\mathbf{x}) = \sum_{i=1}^{n} x_i \ln x_i \) to solve problem (61). (8) displays the reconstructed sparse signal of \( \mathbf{x} \) in the left, the error of \( |f(x^\mu(t)) - f(x^*)| \) in the middle, and the error of \( \|Ax^\mu(t) - b\| \) in the right. All parameters are complied with the requirement. The numerical results are in good accordance with our theoretical results, where the \( |f(x^\mu(t)) - f(x^*)| \) and \( \|Ax^\mu(t) - b\| \) are convergent at the predicted rates.
Example 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:

\[
\min f(x) = \sum_{i=1}^{k} \|x_i\|_1 \\
\text{s.t. } Lx = 0, \quad x \in \mathcal{X},
\]

(62)

where \( L = L_k \otimes I_n \in \mathbb{R}^{kn \times kn} \), \( \mathcal{X} = \{ x \in \mathbb{R}^{kn} | \hat{A}x = b, \hat{A} = \text{blkdiag} \{ A_{m_1 \times n}, \ldots, A_{m_k \times n} \} \} \in \mathbb{R}^{m \times kn}, b \in \mathbb{R}^m \). Note that, the problem (62) 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 \( A_{m_i \times n}, i = 1, \ldots, 5 \) acts as an agent, and 5 agents are connected as a ring. Applying the SADPDMD (50) with \( \nabla \psi^* = P_X \) of affine set \( \mathcal{X} = \{ x \in \mathbb{R}^{kn} | \hat{A}x = b \} \) in (2.4) to solve the problem (62) and all parameters comply with the requirement. The numerical results are displayed in (9). As can be seen from (9) (left) that the variable \( x \) is uniformly convergent and globally asymptotically stable, i.e., \( x_1^\mu = x_2^\mu = \ldots = x_5^\mu \); (8) (middle) shows that the SADPDMD can efficiently solve the problem (62) and reconstruct the original sparse signal; the result in (9) (middle) is in good accordance with our theoretical results, i.e., \( |f(x^\mu(t)) - f(x^*)| \) is convergent at the predicted rates.

![Figure 9: SADPDMD (50) with with \( \alpha = 3 \). (left) Trajectories of \( x \) in 5 agents. (middle) Reconstructed sparse signal of 5 agents. (right) Error of \( |f(x^\mu(t)) - f(x^*)| \).](https://example.com/figure9)

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:

\[
\min f(x) = \sum_{i=1}^{q} \|x_i\|_1 \\
\text{s.t. } \bar{A}x + Ly = b,
\]

(63)

where \( \bar{A} = \text{blkdiag} \{ A_{m_1 \times n_1}, \ldots, A_{m_q \times n_q} \} \in \mathbb{R}^{mq \times n}, \sum_{i=1}^{q} n_q = n \), and \( y \in \mathbb{R}^{mq} \) is an auxiliary variable. Note that, the problem (63) 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 \( \bar{A} \in \mathbb{R}^{100 \times 60} \).
Let every part \( A_{m \times n_i}, i = 1, \ldots, 10 \) be an agent, and 10 agents are connected as a ring. Applying the ADMD (54) to solve the problem (63) and the experimental results are shown in (10). (10) (left) displays the dual variable \( \lambda \) is uniformly convergent and globally asymptotically stable, i.e., \( \lambda_1^\mu = \lambda_2^\mu = \ldots = \lambda_{10}^\mu \); (10) (middle) describes that the SADMD (54) is able to efficiently solve the problem (63) and recover the sparse signal in a distributed way; (10) (right) illustrates \( |f(x^\mu(t)) - f(x^*)| \) of SADMD (54) converges at the predicted rates.

Figure 10: SADMD (54) with \( \alpha = 3 \). (left) Trajectories of \( \lambda \) of 10 agents. (middle) Reconstructed sparse signal. (right) Error of \( |f(x^\mu(t)) - f(x^*)| \).

Acknowledgments and Disclosure of Funding

This work was supported in part by the National Key R&D Program of China (No. 2018AAA0100101), in part by National Natural Science Foundation of China (Grant no. 61932006, 61772434) and in part by the Fundamental Research Funds for the Central Universities (Project No. XDJK2020TY003).

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.

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

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

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.
Hédy Attouch, Zaki Chbani, Jalal M. Fadili, and Hassan Riahi. Fast convergence of dynamical admm via time scaling of damped inertial dynamics. *Journal of Optimization Theory and Applications*, 2021.

Jean-François Aujol, Charles Dossal, and Aude Rondepierre. Convergence rates of the heavy-ball method for quasi-strongly convex optimization. 2020.

Jf Aujol, Charles Dossal, and Aude Rondepierre. Convergence rates of the heavy-ball method under the Lojasiewicz property. *Mathematical Programming*, pages 1–60, 2022.

Aurele Balavoine, Christopher J. Rozell, and Justin K. Romberg. Convergence speed of a dynamical system for sparse recovery. *IEEE Transactions on Signal Processing*, 61:4259–4269, 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.

Heinz H. Bauschke and Patrick L. Combettes. Convex analysis and monotone operator theory in hilbert spaces. In *CMS Books in Mathematics*, 2011.

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*, 25:545–556, 2014.

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

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

Stephen P. Boyd, Neal Parikh, Eric King waih Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2011.

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*, 2021.

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

Alexandre Cabot and Laetitia Paoli. Asymptotics for some vibro-impact problems with a linear dissipation term. *Journal de mathématiques pures et appliquées*, 7:2030037–2030038, 2007.
Xiaojun Chen. Smoothing methods for nonsmooth, nonconvex minimization. *Mathematical Programming*, 134:71–99, 2012.

Jérôme Darbon and Gabriel Provencher 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: Optimization and Control*, 2017.

Mahyar Fazlyab, Alec Koppel, Victor M. Preciado, and Alejandro Ribeiro. A variational approach to dual methods for constrained convex optimization. *2017 American Control Conference (ACC)*, 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.

Chris D. Godsil and Gordon F. Royle. Algebraic graph theory. In *Graduate texts in mathematics*, 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.

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.

John J. Hopfield and David W. Tank. Computing with neural circuits: a model. *Science*, 233 4764:625–33, 1986.

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. *Journal of Machine Learning Research*, 22(17):1–40, 2021.

Walid Krichene, Alexandre M. Bayen, and Peter L. Bartlett. Accelerated mirror descent in continuous and discrete time. In *Conference on Neural Information Processing Systems*, 2015.
Y. Zhao, X. F. Liao, X. He and C. Li

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.

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

Qingshan Liu, Shaofu Yang, and Jun Wang. A collective neurodynamic approach to distributed constrained optimization. *IEEE Transactions on Neural Networks and Learning Systems*, 28:1747–1758, 2017.

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

Yatin Nandwani, Abhishek Pathak, Mausam, and Parag Singla. A primal dual formulation for deep learning with constraints. In *Advances in Neural Information Processing Systems*, 2019.

Yurii 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*, 7:1724–1754, 2014.

Neal Parikh and Stephen P. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1:127–239, 2014.

Boris Polyak. Some methods of speeding up the convergence of iteration methods. *Usr Computational Mathematics and Mathematical Physics*, 4:1–17, 1964.

Guannan Qu and Na Li. On the exponential stability of primal-dual gradient dynamics. *IEEE Control Systems Letters*, 3:43–48, 2019.

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

Weijie J. Su, Stephen P. Boyd, and Emmanuel J. Candès. A differential equation for modeling nesterov’s accelerated gradient method: Theory and insights. In *Journal of Machine Learning Research*, 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.

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:E7351 – E7358, 2016.

Ashia C Wilson, Ben Recht, and Michael I Jordan. A lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, 22:113–1, 2021.
Peng Yi, Yiguang Hong, and Feng Liu. Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. *Automatica*, 74:259–269, 2016.

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 accelerated method with applications to network optimization. *IEEE Transactions on Automatic Control*, pages 1–1, 2022. doi: 10.1109/TAC.2022.3152720.

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. URL http://jmlr.org/papers/v18/16-568.html.

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, Xiao-Fei 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*, PP, 2021.

Zeyuan Allen Zhu and Elad Hazan. Optimal black-box reductions between optimization objectives. In *Conference on Neural Information Processing Systems*, 2016.