Commun. Optim. Theory 2026 (2026) 10 https://doi.org/10.23952/cot.2026.10

Communications in Optimization Theory
Available online at http://cot.mathres.org

GLOBAL AND EXPONENTIAL CONVERGENCE OF A PRIMAL-DUAL DYNAMICAL SYSTEM APPROACH FOR THE SEPARABLE CONVEX OPTIMIZATION PROBLEM

XIN HE[1], RONG HU[2], YA-PING FANG[3,*]

[1]School of Science, Xihua University, Chengdu China
[2]Department of Applied Mathematics, Chengdu University of Information Technology, Chengdu, China
[3]Department of Mathematics, Sichuan University, Chengdu, China

Dedicated to the memory of Professor Hedy Attouch

Abstract. In this paper, we propose a primal-dual dynamical system for the separable convex optimization problem $\min_{x,y} f(x)+g(y), s.t.\, Ax+By=b$. Under the convexity assumptions of f and g, we show that the trajectory of the proposed primal-dual dynamical system globally converges to a saddle point of the problem as the time $t\to+\infty$. When f and/or g satisfy strong convexity and Lipschitz continuity assumptions, along with full rank assumptions on A and/or B, we prove the exponential convergence of the dynamical system approach. Besides, under the metric subregularity condition, we show that the exponential convergence can also be guaranteed without strong convexity and full rank assumptions. Finally, we give numerical results to show the practical performance of the dynamical system approach.

Keywords. Exponential convergence; Global convergence; Primal-dual dynamical system; Separable convex optimization problem.

2020 Mathematics Subject Classification. 34D05, 37N40, 46N10, 90C25.

1. INTRODUCTION

Let $f:\mathbb{R}^n\to\mathbb{R}$ and $g:\mathbb{R}^m\to\mathbb{R}$ be two convex functions. Consider the separable convex optimization problem:

$$\begin{aligned} \min_{x,y} \quad & f(x)+g(y) \\ s.t. \quad & Ax+By=b, \end{aligned} \tag{1.1}$$

where $A\in\mathbb{R}^{p\times n}$, $B\in\mathbb{R}^{p\times m}$, and $b\in\mathbb{R}^p$. This problem plays an important role in diverse applied fields such as, machine learning, signal recovery, structured nonlinear theory and image

*Correesponding author.
E-mail address: ypfang@scu.edu.cn (Y.P. Fang).
Received January 7, 2025; Accepted May 24, 2025.

recovery (see, e.g., [1, 2, 3, 4, 5, 6]). When $g(y) \equiv 0$ and $B \equiv 0$, problem (1.1) becomes the linear equality constrained problem:

$$\min_x f(x), \quad s.t.\, Ax = b. \tag{1.2}$$

When $B \equiv -I_m$ and $b \equiv 0$, where $I_m \in \mathbb{R}^{m\times m}$ is the identity matrix, problem (1.1) becomes the structure problem:

$$\min_x f(x) + g(Ax). \tag{1.3}$$

Primal-dual dynamical systems have been widely used in solving constrained optimization problems (see, e.g., [8, 9, 10, 11, 12, 13, 14, 15, 16, 17]). Here, we mention some dynamical system approaches for the separable problem (1.1), the linear equality constrained problem (1.2) and the structure problem (1.3). When f is twice differentiable, strongly convex, ∇f is Lipschitz continuous, and the matrix A has a full row rank, Qu and Li [18] considered a primal-dual gradient dynamical system with a time constant of the dual part for the problem (1.2), and proved the global exponential stability of the proposed dynamical system. In terms of the augmented Lagrangian function, Tang et al. [19] introduced an augmented primal–dual gradient dynamical system to solve the problem (1.2); When f is strongly convex and has a Lipschitz continuous gradient, and the row vectors of A are linearly independent, they proved the semi-global exponential stability of the proposed dynamical system; They also considered more general constrained optimization problems. Guo et al. [?] proposed a projected dual gradient dynamical system for the problem (1.2), and proved the exponential convergence of the primal residual and dual residual under the assumptions that f is strongly convex and has a Lipschitz continuous gradient; They also considered a projected dual gradient dynamical system for the separable problem (1.1), where f is strongly convex, twice differentiable, g is Lipschitz continuous and strongly convex, and proved that the trajectory of the proposed dynamical system converges exponentially to a KKT point. Considering the continuous limit of ADMM [2], França et al. [20] proposed the following primal dynamical system:

$$\dot{x} + (A^T A)^{-1}\nabla(f(x) + g(Ax)) = 0$$

for solving the problem (1.3), where A has a full column rank, and they proved the $\mathscr{O}(1/t)$ convergence rate under the assumption that f and g are convex. In the Hilbert setting, Boţ et al. [21] proposed a primal-dual dynamical system for the problem (1.3). They showed that the trajectories of the proposed dynamical system weakly converge to a KKT point, and proved the $\mathscr{O}(1/t)$ convergence rate of the ergodic trajectory in general convexity assumptions. Bitterlich et al. [22], Chao and Liu [23] extended the results in [21] to the separable problem (1.1) in different parameter settings. Based on the proximal augmented Lagrangian, Ding and Jovanović [24] proposed a primal-dual gradient dynamical system for the problem (1.3), and demonstrated the global exponential stability of the underlying dynamical system when f is strongly convex and has a Lipschitz continuous gradient, A has a full row rank. Hassan-Moghaddam and Jovanović [25] proposed a proximal gradient flow for the problem (1.3); When $f+g$ satisfies the proximal PL condition, they also proved that the exponential convergence of the objective residual. In the case that f is twice differentiable, strongly convex function and has a Lipschitz continuous gradient, A has a full row rank, Chen and Li [26] considered a primal-dual gradient dynamical system for the problem (1.1); When g is a quadratic function or when g and matrix B together satisfy an inequality condition, the proposed dynamical systems can achieve the global exponential stability. Wang et al. [27] considered a partial primal-dual gradient dynamical system, and showed the exponential stability in the case that f is strongly convex and twice differentiable, g is Lipschitz continuous and strongly convex, A has a full row rank.

In this paper, we propose a primal-dual dynamical system approach for the separable convex optimization problem (1.1). The main contributions are listed as follows:

- We propose a primal-dual dynamical system for the problem (1.1), which can be seen as the continuous counterpart of the proximal ADMM algorithm with a larger dual step size. When f and g are convex, we show that the trajectory of the proposed dynamical system globally converges to a saddle point as $t \to +\infty$.
- When the objective functions f and g, and the matrices A and B satisfy one of cases in Table 1, we prove the exponential convergence of the proposed dynamical system.
- Without strong convexity and full rank assumptions, we show that the exponential convergence can also be guaranteed when the metric subregularity holds.

TABLE 1. Four cases leading to exponential convergence

Case	Strongly convex	Lipschitz continuous	Full row rank	Full column rank
1	f	∇f	A	B
2	f,g	∇f	A	-
3	f	$\nabla f, \nabla g$	$[B,A]$	B
4	f,g	$\nabla f, \nabla g$	$[B,A]$	–

Througout this paper, $\mathbb{R}^n$ denotes the n-dimensional Euclidean space with the scalar product $\langle \cdot,\cdot \rangle$ and the corresponding induced norm $\|\cdot\|$, $\|\cdot\|_1$ denotes the l_1-norm in $\mathbb{R}^n$, and $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix. Given a symmetric matrix M, denote by $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ the largest and smallest eigenvalue of A, respectively. We abuse the notion $\|x\|_M^2 = x^T Mx$ as we allow any symmetric, possibly indefinite, matrix $M \in \mathbb{R}^{n\times n}$. Let us introduce the following partial ordering:

$$M_1 \succcurlyeq M_2 \Longleftrightarrow \|x\|_{M_1}^2 \geq \|x\|_{M_2}^2 \text{ and } M_1 \succ M_2 \Longleftrightarrow \|x\|_{M_1}^2 > \|x\|_{M_2}^2, \quad \forall x \in \mathbb{R}^n \backslash \{0\},$$

where $M_1, M_2 \in \mathbb{R}^{n\times n}$ are two symmetric, possibly indefinite, matrices. Then $M \succ 0$ ($M \succcurlyeq \eta I_n$ with $\eta > 0$) means that M is a symmetric, positive definite matrix. When $M \succ 0$, let $\|x\|_M = \sqrt{x^T Mx}$. Then for any $M \succ 0$, $\lambda_{\min}(M)\|x\|^2 \leq \|x\|_M^2 \leq \lambda_{\max}(M)\|x\|^2$. The notation ∂f denotes the classical convex subdifferential of $f : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\partial f(x) = \{v \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle v, y-x \rangle, \, \forall y \in \mathbb{R}^n\}, \quad \forall x \in \mathbb{R}^n.$$

When f is differentiable, ∂f is equal to its gradient ∇f. The proximal point operator $\mathrm{prox}_{\theta f}$ of a convex function f is defined by

$$\mathrm{prox}_{\theta f}(x) = \arg\min_y \left\{ f(y) + \frac{1}{2\theta}\|y-x\|^2 \right\} = (\theta \partial f + I_n)^{-1},$$

where $\theta > 0$.

The paper is organized as follows: In Section 2, we recall some definitions and lemmas for further analysis. In Section 3, we propose a primal-dual dynamical system for the problem (1.1), and show that the dynamical system is well-defined and admits a unique strong solution under some suitable assumption. In Section 4, we discuss the global convergence of th dynamical system in general convexity assumptions. Section 5 is devoted to the study of the exponential convergence under different conditions. Finally, we give simulation examples to illustrate the performance of the proposed dynamical system in Section 6. Concluding remarks are presented in Section 7, the last section.

2. Preliminaries

In this section, we introduce some standard definitions and lemmas for further analysis.

Definition 2.1. $h:\mathbb{R}^n \to \mathbb{R}$ is ν-convex with $\nu \geq 0$ if

$$h(y) \geq h(x) + \langle s, y-x\rangle + \frac{\nu}{2}\|y-x\|^2,\ \forall x,y \in \mathbb{R}^n,\ s \in \partial h(x).$$

Remark 2.2. When $\nu = 0$, h is a convex function. When $\nu > 0$, h is a strongly convex function with modulus ν. From [28, Example 22.4], if h is ν-convex, then

(i). $h - \frac{\nu}{2}\|\cdot\|^2$ is convex.

(ii). ∂h is ν-monotone, such that

$$\langle s_1 - s_2, x_1 - x_2\rangle \geq \nu\|x_1 - x_2\|^2, \quad \forall s_1 \in \partial h(x_1),\ s_2 \in \partial h(x_2).$$

Definition 2.3. $h:\mathbb{R}^n \to \mathbb{R}$ is η-Lipschitz continuous if, for any $x,y \in \mathbb{R}^n$,

$$\|h(y) - h(x)\| \leq \eta\|y-x\|.$$

Definition 2.4. ([29, Appendix]) Let $T > t_0$. A function $\mathbf{x}:[t_0,T] \to \mathbb{R}^n$ is said to be an absolutely continuous if one of the following equivalent properties holds:

(i). There exists an integrable function $\mathbf{y}:[t_0,T] \to \mathbb{R}^n$ such that

$$\mathbf{x}(t) = \mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{y}(s)ds, \qquad \forall t \in [t_0,T].$$

(ii). For every $\varepsilon > 0$, there exists $\eta > 0$ such that, for any finite family of disjoint intervals $I_k = (a_k,b_k) \subseteq [t_0,T]$ with $\sum_k |b_k - a_k| \leq \eta$, we have $\sum_k \|\mathbf{x}(b_k) - \mathbf{x}(a_k)\| \leq \varepsilon$.

We call a function $\mathbf{x}:[0,\infty) \to \mathbb{R}^n$ is locally absolutely continuous if it is absolutely continuous on every interval $[t_0,T]$.

Remark 2.5. A locally absolutely continuous function $\mathbf{x}$ is differentiable almost everywhere, and its derivative $\dot{\mathbf{x}} = \mathbf{y}$ by integration formula (i).

Definition 2.6. [30, Exercise 3H.4] A set-valued mapping $F:\mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to be metrically subregular at $\hat{x}$ for $\hat{y}$, if $\hat{y} \in F(\hat{x})$ and there exists $\kappa > 0$, along with a neighborhood U of $\hat{x}$, such that

$$dist(x, F^{-1}(\hat{y})) \leq \kappa \cdot dist(\hat{y}, F(x)), \quad \forall x \in U,$$

where $dist(x,\Omega) := \inf\{\|x-y\| | y \in \Omega\}$ for a given subset Ω and vector x in the same space.

Lemma 2.7. [31, Lemma 5.2] *Suppose that* $1 \leq p < +\infty$, $1 \leq q \leq +\infty$, $\mathbf{x}(t) \in L^p([0,+\infty),\mathbb{R}_+)$ *is a locally absolutely continuous function,* $\mathbf{y}(t) \in L^q([0,+\infty),\mathbb{R})$*, and for almost all* t

$$\frac{d}{dt}\mathbf{x}(t) \leq \mathbf{y}(t).$$

Then $\lim_{t\to+\infty}\mathbf{x}(t) = 0$.

Lemma 2.8. *For any $x,\ y\in\mathbb{R}^n$, $\mu>0$, the following inequalities hold.*

$$2\langle x,y\rangle \le \mu\|x\|^2+\frac{1}{\mu}\|y\|^2, \tag{2.1}$$

$$\|x+y\|^2 \ge (1-\frac{1}{\mu})\|x\|^2+(1-\mu)\|y\|^2, \tag{2.2}$$

$$\|x+y\|^2 \le (1+\frac{1}{\mu})\|x\|^2+(1+\mu)\|y\|^2. \tag{2.3}$$

3. PRIMAL-DUAL DYNAMICAL SYSTEM

In this paper, we consider the following primal-dual dynamical system for solving the problem (1.1):

$$\begin{cases} \dot{y}(t)+y(t) &= (\partial g+\rho B^TB+Q)^{-1}(Qy(t)-\rho B^T(Ax(t)-b)-B^T\lambda(t)),\\ \dot{x}(t)+x(t) &= (\partial f+\rho A^TA+P)^{-1}(Px(t)-\rho A^T(B(\dot{y}(t)+y(t))-b)-A^T\lambda(t)),\\ \dot{\lambda}(t) &= \gamma\rho(A(\dot{x}(t)+x(t))+B(\dot{y}(t)+y(t))-b), \end{cases} \tag{3.1}$$

where $(x(t_0),y(t_0),\lambda(t_0))=(x_0,y_0,\lambda_0)$, $t\in[t_0,+\infty)$ and $t_0\ge 0$; $P\in\mathbb{R}^{n\times n}$, $Q\in\mathbb{R}^{m\times m}$ are two symmetric matrices; γ is a larger step size of the dual variable; ∂f, ∂g are the classical convex subdifferentials of f and g, respectively. The first two equalities in (3.1) can be rewritten as

$$\begin{cases} 0\in\partial g(\dot{y}(t)+y(t))+\rho B^TB(\dot{y}(t)+y(t))+Q\dot{y}(t)+\rho B^T(Ax(t)-b)+B^T\lambda(t),\\ 0\in\partial f(\dot{x}(t)+x(t))+\rho A^TA(\dot{x}(t)+x(t))+P\dot{x}(t)+\rho A^T(B(\dot{y}(t)+y(t))-b)+A^T\lambda(t). \end{cases} \tag{3.2}$$

Adopt the following discrete scheme of the dynamical system (3.1): The time step size is fixed to 1, $(\dot{x}(t),\dot{y}(t),\dot{\lambda}(t))\approx(x_{k+1}-x_k,y_{k+1}-y_k,\lambda_{k+1}-\lambda_k)$ and $(x(t),y(t),\lambda(t))\approx(x_k,y_k,\lambda_k)$. Then, we obtain a proximal ADMM with a larger dual step size [32, 33]:

$$\begin{cases} y_{k+1}=\arg\min_{y\in\mathbb{R}^m} g(y)+\frac{\rho}{2}\|By+Ax_k-b+\frac{1}{\rho}\lambda_k\|^2+\frac{1}{2}\|y-y_k\|_Q^2,\\ x_{k+1}=\arg\min_{x\in\mathbb{R}^n} f(x)+\frac{\rho}{2}\|Ax+By_{k+1}-b+\frac{1}{\rho}\lambda_k\|^2+\frac{1}{2}\|x-x_k\|_P^2,\\ \lambda_{k+1}=\lambda_k+\gamma\rho(Ax_{k+1}+By_{k+1}-b). \end{cases}$$

Throughout the paper, we make the following standard assumptions.

Assumption 3.1. f is ν_f-convex and g is ν_g-convex, where $\nu_f,\nu_g\ge 0$. The saddle point set $\Omega\subset\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^p$ of the problem (1.1) is nonempty, such that there exists a saddle point $(x^*,y^*,\lambda^*)\in\Omega$ satisfies the KKT conditions:

$$-A^T\lambda^*\in\partial f(x^*),\quad -B^T\lambda^*\in\partial g(y^*),\quad Ax^*+By^*=b. \tag{3.3}$$

Assumption 3.2. $\rho>0$, $\gamma>0$, $Q\succ 0$ and $P\succ(\frac{1}{4\gamma}-1)\rho A^TA$.

Under Assumption 3.1 and Assumption 3.2, we can obtain the following result.

Lemma 3.3. *Suppose that Assumption 3.1 and Assumption 3.2 hold. Then $(\partial f+\rho A^TA+P)^{-1}$ is a single-valued and $\frac{1}{\eta_f}$- Lipschitz continuous function with $\eta_f=\lambda_{min}(\nu_fI_n+\rho A^TA+P)>0$, and $(\partial g+\rho B^TB+Q)^{-1}$ is single-valued and $\frac{1}{\eta_g}$- Lipschitz continuous function with $\eta_g=\lambda_{min}(\nu_gI_m+\rho B^TB+Q)>0$.*

Proof. For notion simplicity, denote $S(x) = (\partial f + \rho A^T A + P)^{-1}(x)$. Define $h : \mathbb{R}^n \to \mathbb{R}$ by

$$h(x) = f(x) - \frac{\nu_f}{2}\|x\|^2 + \frac{1}{2}\|x\|^2_{\nu_f I_n + \rho A^T A + P - \eta_f I_n},$$

where $\eta_f = \lambda_{\min}(\nu_f I_n + \rho A^T A + P)$. From Assumption 3.2 and Remark 2.2, we know that $\eta_f > 0$ and h is a convex function, and then $\partial h = \partial f + \rho A^T A + P - \eta_f I_n$ is maximal monotone. Since

$$\begin{aligned} S &= (\partial f + \rho A^T A + P - \eta_f I_n + \eta_f I_n)^{-1} \\ &= \frac{1}{\eta_f}(I_n + \frac{1}{\eta_f}(\partial f + \rho A^T A + P - \eta_f I_n))^{-1} \\ &= \frac{1}{\eta_f}(I_n + \frac{1}{\eta_f}\partial h)^{-1}, \end{aligned}$$

it follows from [34, Proposition 3.4] that $S = (\partial f + \rho A^T A + P)^{-1}$ is single-valued. Since h is convex, $S^{-1} = \partial f + \rho A^T A + P = \partial h + \eta_f I_n$ is η_f- monotone. It follows that

$$\langle x - y, S(x) - S(y)\rangle \geq \eta_f \|S(x) - S(y)\|^2.$$

This yields $\|S(x) - S(y)\| \leq \frac{1}{\eta_f}\|x - y\|$. So $(\partial f + \rho A^T A + P)^{-1}$ is $\frac{1}{\eta_f}$- Lipschitz continuous.

Similarly, we can prove that $(\partial g + \rho B^T B + Q)^{-1}$ is single-valued and $\frac{1}{\eta_g}$- Lipschitz continuous with $\eta_g = \lambda_{\min}(\nu_g I_m + \rho B^T B + Q) > 0$. □

By Lemma 3.3 and the Cauchy-Lipschitz-Picard Theorem, we can obtain the following result.

Proposition 3.4. *Suppose that Assumption 3.1 and Assumption 3.2 hold. Then the primal-dual dynamical system (3.1) is well-defined. Moreover, for any initial point $(x_0, y_0, \lambda_0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$, the dynamical system (3.1) admits a unique strong solution $(x, y, \lambda) : [t_0, +\infty) \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$, such that*

(i). $(x(t_0), y(t_0), \lambda(t_0)) = (x_0, y_0, \lambda_0)$.
(ii). *The trajectory $(x(t), y(t), \lambda(t))$ are locally absolutely continuous, and then $(x(t), y(t), \lambda(t))$ are differentiable almost everywhere.*
(iii). *For almost every $t \in [t_0, +\infty)$, (3.1) is satisfied.*

4. GLOBAL CONVERGENCE

In this section, we will investigate the global convergence of the dynamical system (3.1) in general convex assumptions.

Lemma 4.1. *Suppose that Assumption 3.1 and Assumption 3.2 hold. Let $(x(t), y(t), \lambda(t))$ be the strong solution of the dynamical system (3.1) with an initial point (x_0, y_0, λ_0). Then $(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t))$ is locally absolutely continuous, $(\ddot{x}(t), \ddot{y}(t), \ddot{\lambda}(t))$ exists almost everywhere, and the following inequality holds for almost every $t \in [t_0 + \infty)$ with some $L > 0$:*

$$\|\ddot{x}(t)\| + \|\ddot{y}(t)\| + \|\ddot{\lambda}(t)\| \leq L(\|\dot{x}(t)\| + \|\dot{y}(t)\| + \|\dot{\lambda}(t)\|).$$

Proof. Let $T > t_0$ and $s,t \in [t_0,T]$. By Lemma 3.3, (3.1) and the triangle inequality we have

$$\begin{aligned}
&\|\dot{y}(s)+y(s)-(\dot{y}(t)+y(t))\| \\
&\le \frac{1}{\eta_g}\|Q(y(s)-y(t))-\rho B^T A(x(s)-x(t))-B^T(\lambda(s)-\lambda(t))\| \\
&\le L_1(\|x(s)-x(t)\|+\|y(s)-y(t)\|+\|\lambda(s)-\lambda(t)\|),
\end{aligned} \tag{4.1}$$

where $L_1 = \frac{1}{\eta_g}\max\{\|Q\|,\rho\|B^TA\|,\|B^T\|\} > 0$. Similarly, there exist $L_2,\ L_3 > 0$ such that

$$\begin{aligned}
&\|\dot{x}(s)+x(s)-(\dot{x}(t)+x(t))\| \\
&\quad \le L_2(\|x(s)-x(t)\|+\|y(s)-y(t)\|+\|\lambda(s)-\lambda(t)\|)
\end{aligned} \tag{4.2}$$

and

$$\|\dot{\lambda}(s)-\dot{\lambda}(t)\| \le L_3(\|x(s)-x(t)\|+\|y(s)-y(t)\|+\|\lambda(s)-\lambda(t)\|). \tag{4.3}$$

Adding (4.1)-(4.3), by the triangle inequality we have

$$\begin{aligned}
&\|\dot{x}(s)-\dot{x}(t)\|+\|\dot{y}(s)-\dot{y}(t)\|+\|\dot{\lambda}(s)-\dot{\lambda}(t)\| \\
&\quad \le L(\|x(s)-x(t)\|+\|y(s)-y(t)\|+\|\lambda(s)-\lambda(t)\|),
\end{aligned} \tag{4.4}$$

where $L = L_1+L_2+L_3+2$.

Since $(x(t),y(t),\lambda(t))$ are locally absolutely continuous, for any $\varepsilon > 0$, there exists $\eta > 0$ such that, for any finite family of disjoint intervals $I_k = (a_k,b_k) \subseteq [t_0,T]$ with $\sum_k |b_k-a_k| \le \eta$, we have

$$\sum_k \|x(b_k)-x(a_k)\| \le \frac{\varepsilon}{3L}, \quad \sum_k \|y(b_k)-y(a_k)\| \le \frac{\varepsilon}{3L}, \quad \sum_k \|\lambda(b_k)-\lambda(a_k)\| \le \frac{\varepsilon}{3L}.$$

Consequently, from (4.4), we get

$$\sum_k \|\dot{x}(b_k)-\dot{x}(a_k)\| \le \varepsilon, \quad \sum_k \|\dot{y}(b_k)-\dot{y}(a_k)\| \le \varepsilon, \quad \sum_k \|\dot{\lambda}(b_k)-\dot{\lambda}(a_k)\| \le \varepsilon.$$

So $(\dot{x}(t),\dot{y}(t),\dot{\lambda}(t))$ is locally absolutely continuous. By Remark 2.5, $(\ddot{x}(t),\ddot{y}(t),\ddot{\lambda}(t))$ exists almost everywhere. Taking $s = t+h$ with $h > 0$ in (4.4), dividing both sides of (4.4) by h and letting $h \to 0$, we have

$$\|\ddot{x}(t)\|+\|\ddot{y}(t)\|+\|\ddot{\lambda}(t)\| \le L(\|\dot{x}(t))\|+\|\dot{y}(t)\|+\|\dot{\lambda}(t)\|).$$

This inequality holds for almost every $t \in [t_0,+\infty)$ since $T > t_0$ is arbitrary. □

For notation simplicity, we introduce

$$G_0 = \begin{pmatrix} I_n & & \\ & I_m & \\ & & \gamma I_p \end{pmatrix}, \quad G_1 = \begin{pmatrix} \hat{P} & & \\ & Q & \\ & & \frac{1}{\rho} I_p \end{pmatrix}, \quad G = G_0^{-1}G_1 = \begin{pmatrix} \hat{P} & & \\ & Q & \\ & & \frac{1}{\rho\gamma} I_p \end{pmatrix}$$

with $\hat{P} = P+\rho A^TA$. From Assumption (3.2), we have $G_0 \succ 0$, $G_1 \succ 0$, and $G \succ 0$. Let $(x(t),y(t),\lambda(t))$ be a strong solution of the dynamical system (3.1). Denote

$$\bar{\lambda}(t) = \lambda(t)+\frac{1}{\gamma}\dot{\lambda}(t),$$

$$u(t)=\begin{pmatrix}x(t)\\ y(t)\\ \lambda(t)\end{pmatrix},\quad u^*=\begin{pmatrix}x^*\\ y^*\\ \lambda^*\end{pmatrix},\quad \bar{u}(t)=\begin{pmatrix}x(t)+\dot{x}(t)\\ y(t)+\dot{y}(t)\\ \bar{\lambda}(t)\end{pmatrix}.$$

where $u^*\in\Omega$ is a saddle point. It is easy to verify that

$$\dot{u}(t)=G_0(\bar{u}(t)-u(t)). \tag{4.5}$$

Lemma 4.2. *Suppose that Assumption 3.1 and Assumption 3.2 hold. Define the energy function* $\mathscr{E}(t)=\frac{1}{2}\|u(t)-u^*\|_G^2$. *Then*

$$\begin{aligned}\dot{\mathscr{E}}(t) \quad &\leq \quad \frac{1}{\gamma}\langle A\dot{x}(t),\dot{\lambda}(t)\rangle-\|\bar{u}(t)-u(t)\|_{G_1}^2 \\ &\qquad -\nu_f\|\dot{x}(t)+x(t)-x^*\|^2-\nu_g\|\dot{y}(t)+y(t)-y^*\|^2.\end{aligned} \tag{4.6}$$

Proof. Computing the time derivative of $\mathscr{E}(t)$, we have

$$\begin{aligned}\dot{\mathscr{E}}(t) &= \langle\dot{u}(t),G(u(t)-u^*)\rangle \\ &= \langle\bar{u}(t)-u(t),G_1(u(t)-u^*)\rangle \\ &= \langle\bar{u}(t)-u(t),G_1(\bar{u}(t)-u^*)\rangle-\|\bar{u}(t)-u(t)\|_{G_1}^2,\end{aligned} \tag{4.7}$$

where the second equality follows from (4.5). It follows from (3.1) and (3.2) that

$$-B^T\bar{\lambda}(t)-Q\dot{y}(t)+\rho B^TA\dot{x}(t)\in\partial g(\dot{y}(t)+y(t)) \tag{4.8}$$

and

$$-A^T\bar{\lambda}(t)-P\dot{x}(t)\in\partial f(\dot{x}(t)+x(t)). \tag{4.9}$$

Since $u^*=(x^*,y^*,\lambda^*)$ is a saddle point, from (3.3) and (3.1) we have

$$\frac{1}{\gamma\rho}\dot{\lambda}(t)=A(\dot{x}(t)+x(t)-x^*)+B(\dot{y}(t)+y(t)-y^*). \tag{4.10}$$

Then, we can compute

$$\begin{aligned}&\langle\bar{u}(t)-u(t),G_1(\bar{u}(t)-u^*)\rangle \\ &=\langle\hat{P}\dot{x}(t),\dot{x}(t)+x(t)-x^*\rangle+\langle Q\dot{y}(t),\dot{y}(t)+y(t)-y^*\rangle+\langle\frac{1}{\rho\gamma}\dot{\lambda}(t),\bar{\lambda}(t)-\lambda^*\rangle \\ &\overset{(4.10)}{=}\langle P\dot{x}(t)+A^T\bar{\lambda}(t)-A^T\lambda^*,\dot{x}(t)+x(t)-x^*\rangle+\frac{1}{\gamma}\langle A\dot{x}(t),\dot{\lambda}(t)\rangle \\ &\quad+\langle Q\dot{y}(t)+B^T\bar{\lambda}(t)-\rho B^TA\dot{x}(t)-B^T\lambda^*,(\dot{y}(t)+y(t)-y^*)\rangle \\ &\leq\frac{1}{\gamma}\langle A\dot{x}(t),\dot{\lambda}(t)\rangle-\nu_f\|\dot{x}(t)+x(t)-x^*\|^2-\nu_g\|\dot{y}(t)+y(t)-y^*\|^2,\end{aligned}$$

where the last inequality follows from (4.8), (4.9) and Assumption 3.1. This together with (4.7) yields the desired result. □

Remark 4.3. Since $\nu_f\geq 0$ and $\nu_g\geq 0$, from (4.6) we have

$$\dot{\mathscr{E}}(t)\leq\frac{1}{\gamma}\langle A\dot{x}(t),\dot{\lambda}(t)\rangle-\|\bar{u}(t)-u(t)\|_{G_1}^2. \tag{4.11}$$

In fact, the above inequality still holds when f and g are convex functions (in this case, $\nu_f=\nu_g=0$).

Theorem 4.4. *Suppose that Assumption 3.1 and Assumption 3.2 hold. Let $u(t) = (x(t), y(t), \lambda(t))$ be a strong solution of the dynamical system (3.1) with an initial point (x_0, y_0, λ_0). Then the following statements are true:*

(i). *For every $u^* = (x^*, y^*, \lambda^*) \in \Omega$, $\lim_{t\to+\infty} \mathscr{E}(t)$ exists.*
(ii). *$(\dot{x}(t), \dot{y}(t), \dot{\lambda}(t)) \to 0$ as $t \to +\infty$.*
(iii). *There exists a saddle point $u^* = (x^*, y^*, \lambda^*)$ such that $u(t) \to u^*$ as $t \to +\infty$. As a consequence, $(x(t), y(t), \lambda(t)) \to (x^*, y^*, \lambda^*)$ as $t \to +\infty$.*

Proof. (i). Since $P \succ (\frac{1}{4\gamma} - 1)\rho A^T A$, there exists $\varepsilon > 0$ such that $P + \rho A^T A \succcurlyeq \frac{\rho}{4\gamma} A^T A + \varepsilon I_n$. On the other hand, there exists $\beta_0 > 0$ such that $\frac{\varepsilon}{2} I_n \succcurlyeq \beta_0 A^T A$. Taking $\beta = \frac{\rho}{4\gamma} + \beta_0$, then

$$P + (\rho - \beta)A^T A \succcurlyeq \frac{\varepsilon}{2} I_n + (\frac{\varepsilon}{2} I_n - \beta_0 A^T A) \succ 0 \tag{4.12}$$

and

$$\frac{1}{\rho} - \frac{1}{4\beta\gamma} = \frac{1}{\rho} - \frac{1}{\rho + 4\gamma\beta_0} > 0. \tag{4.13}$$

Following from (2.1), we have

$$\frac{1}{\gamma}\langle A\dot{x}(t), \dot{\lambda}(t)\rangle \leq \beta\|A\dot{x}(t)\|^2 + \frac{1}{4\gamma^2\beta}\|\dot{\lambda}(t)\|^2 = \|\dot{x}(t)\|^2_{\beta A^T A} + \frac{1}{4\gamma^2\beta}\|\dot{\lambda}(t)\|^2. \tag{4.14}$$

By the definition of $\bar{u}(t)$, $u(t)$, G_1, it follows from (4.11) and (4.14) that

$$\begin{aligned} \dot{\mathscr{E}}(t) &\leq \frac{1}{\gamma}\langle A\dot{x}(t), \dot{\lambda}(t)\rangle - \frac{1}{\rho\gamma}\|\dot{\lambda}(t)\|^2 - \|\dot{x}(t)\|^2_{P+\rho A^T A} - \|\dot{y}(t)\|^2_Q \\ &\leq -\|\dot{x}(t)\|^2_{P+(\rho-\beta)A^T A} + \frac{1}{\gamma}(\frac{1}{4\gamma\beta} - \frac{1}{\rho})\|\dot{\lambda}(t)\|^2 - \lambda_{\min}(Q)\|\dot{y}(t)\|^2. \end{aligned} \tag{4.15}$$

Since $Q \succ 0$ and $\lambda_{min}(Q) > 0$, this together with (4.12) and (4.13) implies $\dot{\mathscr{E}}(t) \leq 0$. Since $G \succ 0$, $\mathscr{E}(t)$ is a nonincreasing and nonnegative function, this yields (i).

(ii). Integrating (4.15) on the interval $[t_0, +\infty]$, using (4.12), (4.13) and $Q \succ 0$, form (i) we can get

$$\int_{t_0}^{+\infty} \|\dot{x}(t)\|^2 < +\infty, \int_{t_0}^{+\infty} \|\dot{y}(t)\|^2 < +\infty, \int_{t_0}^{+\infty} \|\dot{z}(t)\|^2 < +\infty,$$

which implies

$$\dot{x}(\cdot) \in L^2([t_0, +\infty), \mathbb{R}^n),\ \dot{y}(\cdot) \in L^2([t_0, +\infty), \mathbb{R}^m),\ \dot{\lambda}(\cdot) \in L^2([t_0, +\infty), \mathbb{R}^p).$$

This together with Lemma 4.1 yields

$$\ddot{x}(\cdot) \in L^2([0, +\infty), \mathbb{R}^n),\ \ddot{y}(\cdot) \in L^2([0, +\infty), \mathbb{R}^m),\ \ddot{\lambda}(\cdot) \in L^2([0, +\infty), \mathbb{R}^p).$$

Then, for almost every $t \in [t_0, +\infty)$,

$$\frac{d}{dt}\|\dot{x}(t)\|^2 = 2\langle \ddot{x}(t), \dot{x}(t)\rangle \leq \|\ddot{x}(t)\|^2 + \|\dot{x}(t)\|^2 \in L^1([t_0, +\infty), \mathbb{R}).$$

Applying Lemma 2.7, we have

$$\dot{x}(t) \to 0 \quad \text{as } t \to +\infty.$$

Similarly, we get $\dot{y}(t) \to 0$, $\dot{\lambda}(t) \to 0$ as $t \to +\infty$, this is (ii).

(iii). Let $\hat{u}=(\hat{x},\hat{y},\hat{\lambda})$ be any sequential cluster of $u(t)$ with $u(t_n)=(x(t_n),y(t_n),\lambda(t_n))\to\hat{u}$ as $t_n\to+\infty$. From the third equality of (3.1) and (ii) we get

$$\lim_{t_n\to+\infty} Ax(t_n)+By(t_n)-b=0. \tag{4.16}$$

It follows from (4.8) and (4.9) that

$$-B^T\bar{\lambda}(t_n)-Q\dot{y}(t_n)+\rho B^TA\dot{x}(t_n)\in\partial g(\dot{y}(t_n)+y(t_n))$$

and

$$-A^T\bar{\lambda}(t_n)-P\dot{x}(t_n)\in\partial f(\dot{x}(t_n)+x(t_n))$$

Let $t_n\to+\infty$. By (ii) and [28, Proposition 20.38] we get

$$-B^T\hat{\lambda}\in\partial g(\hat{y}),\ -A^T\hat{\lambda}\in\partial f(\hat{x}).$$

This together with (4.16) implies $(\hat{x},\hat{y},\hat{\lambda})$ is a saddle point. So every sequential cluster point of $u(t)$ belongs to Ω. By (i) and $G\succ 0$, $u(t)$ is bounded on $[t_0,+\infty)$. Let u_1, u_2 be two sequential cluster points, say $u(t_n^1)\to u_1$ and $u(t_n^2)\to u_2$. Since u_1 and u_2 belong to Ω, from (i), $\|u(t)-u_1\|_G$ and $\|u(t)-u_2\|_G$ converge as $t\to\infty$. In turn

$$2\langle u(t),G(u_1-u_2)\rangle=\|u(t)-u_2\|_G^2-\|u(t)-u_1\|_G^2+\|u_1\|^2-\|u_2\|_G^2,$$

so, $\langle u(t),G(u_1-u_2)\rangle$ converges to a constant C. Passing to the limit along $u(t_n^1)$ and along $u(t_n^2)$ yields

$$C=\langle u_1,G(u_1-u_2)\rangle=\langle u_2,G(u_1-u_2)\rangle,$$

and then $\|u_1-u_2\|_G^2=0$, this together with $G\succ 0$ implies $u_1=u_2$. So $u(t)$ has a unique sequential cluster point in Ω, there exists a saddle point $u^*=(x^*,y^*,\lambda^*)$ such that $u(t)\to u^*$ as $t\to+\infty$. As a consequence, $(x(t),y(t),\lambda(t))\to(x^*,y^*,\lambda^*)$ as $t\to+\infty$. $\square$

Remark 4.5. It is worth noting that the proof of Theorem 2 only requires the inequality (4.11). Therefore, by assuming only the convexity of the objective functions f and g, the global convergence of the dynamical system (3.1) can be guaranteed.

Remark 4.6. Taking $Q=\theta_1 I_m-\rho B^TB$ with $\theta_1>\rho\|B^TB\|$, $P=\theta_2 I_n-\rho A^TA$ with $\theta_2>\frac{\rho}{4\gamma}\|A^TA\|$, then Assumption 3.2 holds. In this case, the dynamical system (3.1) collapses to the following proximal primal-dual dynamical system:

$$\begin{cases}\dot{y}(t)+y(t) & =\operatorname{prox}_{\frac{1}{\theta_1}g}(y(t)-\frac{1}{\theta_1}p(t)),\\ \dot{x}(t)+x(t) & =\operatorname{prox}_{\frac{1}{\theta_2}f}(x(t)-\frac{1}{\theta_2}q(t)),\\ \dot{\lambda}(t) & =\gamma\rho(A(\dot{x}(t)+x(t))+B(\dot{y}(t)+y(t))-b),\end{cases}$$

where

$$p(t)=B^T(\rho(Ax(t)+By(t)-b)+\lambda(t))$$

and

$$q(t)=A^T(\rho(Ax(t)+B(y(t)+\dot{y}(t))-b)+\lambda(t)).$$

This can be seen as the continuous counterpart of the linearized ADMM algorithm.

5. Exponential convergence

As shown by Proposition 3.4 and Theorem 4.4, under mild conditions, the dynamical system (3.1) with an initial point admits a unique strong solution which converges to a saddle point of the problem (1.1) as the time $t \to +\infty$. In this section, we always assume that $u(t) = (x(t), y(t), \lambda(t))$ is the unique strong solution of the dynamical system (3.1) with an initial point $u(t_0) = (x_0, y_0, \lambda_0)$ and $u(t))$ converges to $u^* = (x^*, y^*, \lambda^*) \in \Omega$ as $t \to +\infty$.

Firstly, we show the exponential convergence of the dynamical system (3.1) under some strongly convex and Lipschitz continuity assumptions on f and g, along with certain rank assumptions on A and B.

Lemma 5.1. *Suppose that B has a full column rank. Then for any $\mu_1 > 0$,*

$$\|\dot{y}(t) + y(t) - y^*\|^2 \le c_1 \|\dot{x}(t) + x(t) - x^*\|^2 + c_2 \|\dot{\lambda}(t)\|^2,$$

where $c_1 = \lambda_{\min}^{-1}(B^T B)(1+\mu_1)\|A^T A\| > 0$, $c_2 = \lambda_{\min}^{-1}(B^T B)(1+\frac{1}{\mu_1})\frac{1}{\rho^2\gamma^2} > 0$.

Proof. It follows from (4.10) and (2.3) that for any $\mu_1 > 0$,

$$\begin{aligned} \|B(\dot{y}(t) + y(t) - y^*)\|^2 &= \left\| \frac{1}{\rho\gamma}\dot{\lambda}(t) - A(\dot{x}(t) + x(t) - x^*) \right\|^2 \\ &\le (1+\frac{1}{\mu_1})\frac{1}{\rho^2\gamma^2}\|\dot{\lambda}(t)\|^2 + (1+\mu_1)\|A^T A\|\|\dot{x}(t) + x(t) - x^*\|^2. \end{aligned} \tag{5.1}$$

Since B has a full column rank,

$$\|B(\dot{y}(t) + y(t) - y^*)\|^2 \ge \lambda_{\min}(B^T B)\|\dot{y}(t) + y(t) - y^*\|^2.$$

This together with (5.1) yields the result. □

Lemma 5.2. *Suppose that f is L_f-Lipschitz continuous and A has a full row rank. Then for any $\mu_2 > 1$,*

$$\|\bar{\lambda}(t) - \lambda^*\|^2 \le c_3 \|\dot{x}(t) + x(t) - x^*\|^2 + c_4 \|\dot{x}(t)\|^2,$$

where $c_3 = \lambda_{\min}^{-1}(AA^T)(1-\frac{1}{\mu_2})^{-1}L_f^2 > 0$, $c_4 = \lambda_{\min}^{-1}(AA^T)\mu_2\|P\|^2 > 0$.

Proof. Since ∇f is Lipschitz continuous and $-A^T\lambda^* = \nabla f(x^*)$, it follows from (4.9) that

$$\begin{aligned} \|-A^T\bar{\lambda}(t) - P\dot{x}(t) + A^T\lambda^*\|^2 &= \|\nabla f(\dot{x}(t) + x(t)) - \nabla f(x^*)\|^2 \\ &\le L_f^2 \|\dot{x}(t) + x(t) - x^*\|^2. \end{aligned}$$

This together with (2.2) implies

$$(1-\frac{1}{\mu_2})\|A^T(\bar{\lambda}(t) - \lambda^*)\|^2 \le L_f^2\|\dot{x}(t) + x(t) - x^*\|^2 + (\mu_2 - 1)\|P\|^2\|\dot{x}(t)\|^2$$

for any $\mu_2 > 1$. Since A has a full row rank,

$$\|A^T(\bar{\lambda}(t) - \lambda^*)\|^2 \ge \lambda_{\min}(AA^T)\|\bar{\lambda}(t) - \lambda^*\|^2$$

with $\lambda_{\min}(AA^T) > 0$. The conclusion follows from the last two inequalities. □

Lemma 5.3. *Suppose that ∇f is L_f-Lipschitz continuous, and ∇g is L_g-Lipschitz continuous, $[B,A]$ has a full row rank. Then for any $\mu_3 > 1$ and $\mu_4 > 0$,*

$$\|\bar{\lambda}(t)-\lambda^*\|^2 \leq c_5\|\dot{x}(t)+x(t)-x^*\|^2 + c_6\|\dot{y}(t)+y(t)-y^*\|^2 + c_7\|\dot{x}(t)\|^2 + c_8\|\dot{y}(t)\|^2,$$

where $\hat{c} = \lambda_{\min}^{-1}([B,A][B,A]^T) > 0$, $c_5 = (1-\frac{1}{\mu_3})^{-1}\hat{c}L_f^2 > 0$, $c_6 = (1-\frac{1}{\mu_3})^{-1}\hat{c}L_g^2 > 0$, $c_7 = \hat{c}\mu_3(1+\mu_4)\|[\rho A^T B, P^T]\|^2 > 0$ and $c_8 = \hat{c}\mu_3(1+\frac{1}{\mu_4})\|Q\|^2 > 0$.

Proof. Since $-A^T\lambda^* = \nabla f(x^*)$, $-B^T\lambda^* = \nabla g(y^*)$ and ∇f and ∇g are Lipschitz continuous, from (4.8) and (4.9) we get

$$\begin{aligned}
&\left\| \begin{bmatrix} -B^T \\ -A^T \end{bmatrix} (\bar{\lambda}(t)-\lambda^*) + \begin{bmatrix} -Q \\ 0 \end{bmatrix} \dot{y}(t) + \begin{bmatrix} \rho B^T A \\ P \end{bmatrix} \dot{x}(t) \right\|^2 \\
&= \|\nabla g(\dot{y}(t)+y(t)) - \nabla g(y^*)\|^2 + \|\nabla f(\dot{x}(t)+x(t)) - \nabla f(x^*)\|^2 \\
&\leq L_g^2\|\dot{y}(t)+y(t)-y^*\|^2 + L_f^2\|\dot{x}(t)+x(t)-x^*\|^2.
\end{aligned} \tag{5.2}$$

Applying (2.2) we have

$$\begin{aligned}
&\left\| \begin{bmatrix} -B^T \\ -A^T \end{bmatrix} (\bar{\lambda}(t)-\lambda^*) + \begin{bmatrix} -Q \\ 0 \end{bmatrix} \dot{y}(t) + \begin{bmatrix} \rho B^T A \\ P \end{bmatrix} \dot{x}(t) \right\|^2 \\
&\geq (1-\frac{1}{\mu_3}) \left\| \begin{bmatrix} -B^T \\ -A^T \end{bmatrix} (\bar{\lambda}(t)-\lambda^*) \right\|^2 \\
&\quad +(1-\mu_3) \left\| \begin{bmatrix} -Q \\ 0 \end{bmatrix} \dot{y}(t) + \begin{bmatrix} \rho B^T A \\ P \end{bmatrix} \dot{x}(t) \right\|^2
\end{aligned} \tag{5.3}$$

for any $\mu_3 > 1$. Applying (2.3) we have

$$\begin{aligned}
&\left\| \begin{bmatrix} -Q \\ 0 \end{bmatrix} \dot{y}(t) + \begin{bmatrix} \rho B^T A \\ P \end{bmatrix} \dot{x}(t) \right\|^2 \\
&\leq (1+\frac{1}{\mu_4})\|Q\|^2\|\dot{y}(t)\|^2 + (1+\mu_4)\|[\rho A^T B, P^T]\|^2\|\dot{x}(t)\|^2
\end{aligned} \tag{5.4}$$

for any $\mu_4 > 0$. Combining (5.2)-(5.4) and using the fact $[B,A]$ has a full row rank, we obtain the desired result. □

With the above three lemmas in hand, we prove the exponential convergence of the dynamical system (3.1).

Theorem 5.4. *Suppose that Assumption 3.1 and Assumption 3.2 hold, ∇f is L_f-Lipschitz continuous, and $\nu_f > 0$, and one of the following conditions is satisfied:*

(i). *A has a full row rank and B has a full column rank.*
(ii). *A has a full row rank and $\nu_g > 0$.*
(iii). *∇g is L_g-Lipschitz continuous, B has a full column rank, $[B,A]$ has a full row rank.*
(iv). *∇g is L_g-Lipschitz continuous, $\nu_g > 0$, $[B,A]$ has a full row rank.*

Then, there exists $\sigma > 0$ such that for almost every $t \geq t_0$,

$$\|u(t)-u^*\|_G \leq e^{-\sigma(t-t_0)}\|u(t_0)-u^*\|_G.$$

Proof. By (4.6) and (4.14), there exists $\beta > 0$ with $P + (\rho - \beta)A^T A \succ 0$ and $\frac{1}{\rho} - \frac{1}{4\gamma\beta} > 0$ such that

$$\begin{aligned}
\dot{\mathscr{E}}(t) &\leq -\|\dot{x}(t)\|^2_{P+(\rho-\beta)A^TA} - \frac{1}{\gamma}(\frac{1}{\rho} - \frac{1}{4\gamma\beta})\|\dot{\lambda}(t)\|^2 - \|\dot{y}(t)\|^2_Q \\
&\quad -\nu_f\|\dot{x}(t) + x(t) - x^*\|^2 - \nu_g\|\dot{y}(t) + y(t) - y^*\|^2 \\
&\leq -C_1(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2 + \|\dot{\lambda}(t)\|^2) \\
&\quad -\nu_f\|\dot{x}(t) + x(t) - x^*\|^2 - \nu_g\|\dot{y}(t) + y(t) - y^*\|^2,
\end{aligned} \tag{5.5}$$

where $C_1 = \min\{\lambda_{\min}(P + (\rho-\beta)A^TA), \frac{1}{\gamma}(\frac{1}{\rho} - \frac{1}{4\gamma\beta}), \lambda_{\min}(Q)\} > 0$.

Following from the inequality (2.3) with $\mu = \frac{1}{2}$ and the definition of $\mathscr{E}(t)$, we can compute

$$\begin{aligned}
\mathscr{E}(t) &= \frac{1}{2}\|x(t) - x^*\|^2_{\hat{P}} + \frac{1}{2}\|y(t) - y^*\|^2_Q + \frac{1}{2\rho\gamma}\|\lambda(t) - \lambda^*\|^2 \\
&\leq \lambda_{\max}(\hat{P})(\|\dot{x}(t) + x(t) - x^*\|^2 + \|\dot{x}(t)\|^2) \\
&\quad \lambda_{\max}(Q)(\|\dot{y}(t) + y(t) - y^*\|^2 + \|\dot{y}(t)\|^2) \\
&\quad + \frac{1}{\rho\gamma}\left(\left\|\frac{1}{\gamma}\dot{\lambda}(t) + \lambda(t) - \lambda^*\right\|^2 + \frac{1}{\gamma^2}\|\dot{\lambda}(t)\|^2\right) \\
&\leq C_2(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2 + \|\dot{\lambda}(t)\|^2) + \lambda_{\max}(\hat{P})\|\dot{x}(t) + x(t) - x^*\|^2 \\
&\quad + \lambda_{\max}(Q)\|\dot{y}(t) + y(t) - y^*\|^2 + \frac{1}{\rho\gamma}\|\bar{\lambda}(t) - \lambda^*\|^2,
\end{aligned} \tag{5.6}$$

where $C_1 = \max\{\lambda_{\max}(\hat{P}), \lambda_{\max}(Q), \frac{1}{\rho\gamma^3}\} > 0$.

If there exists $\sigma > 0$ such that

$$\dot{\mathscr{E}}(t) + \sigma\mathscr{E}(t) \leq 0, \tag{5.7}$$

then multiplying by $e^{\sigma t}$ and integrating over $[t_0, t]$ we have

$$e^{\sigma t}\mathscr{E}(t) \leq e^{\sigma t_0}\mathscr{E}(t_0).$$

This together with the definition of $\mathscr{E}(t)$ implies

$$\|u(t) - u^*\|_G \leq e^{-\sigma(t-t_0)}\|u(t_0) - u^*\|_G.$$

To complete the proof, it suffices to show that the inequality (5.7) holds when one of the conditions $(i) - (iv)$ is satisfied.

(i). It follows from Lemma 5.1 and Lemma 5.2 that

$$\begin{aligned}
&\lambda_{\max}(Q)\|\dot{y}(t) + y(t) - y^*\|^2 + \frac{1}{\rho\gamma}\|\bar{\lambda}(t) - \lambda^*\|^2 \\
&\leq (c_1\lambda_{\max}(Q) + \frac{c_3}{\rho\gamma})\|\dot{x}(t) + x(t) - x^*\|^2 + c_2\lambda_{\max}(Q)\|\dot{\lambda}(t)\|^2 + \frac{c_4}{\rho\gamma}\|\dot{x}(t)\|^2,
\end{aligned}$$

which together with (5.5) and (5.6) implies that (5.7) holds with

$$\sigma = \max\left\{C_1(C_2 + c_2\lambda_{\max}(Q) + \frac{c_4}{\rho\gamma})^{-1}, \nu_f(\lambda_{\max}(\hat{P}) + c_1\lambda_{\max}(Q) + \frac{c_3}{\rho\gamma})^{-1}\right\} > 0.$$

(ii). It follows from Lemma 5.2, (5.5) and (5.6) that (5.7) holds with

$$\sigma = \max\left\{C_1(C_2+\frac{c_4}{\rho\gamma})^{-1}, \nu_f(\lambda_{\max}(\hat{P})+\frac{c_3}{\rho\gamma})^{-1}, \nu_g\lambda_{\max}(Q)^{-1}\right\} > 0.$$

(iii). It follows from Lemma 5.1 and Lemma 5.3 that

$$\begin{aligned}
&\lambda_{\max}(Q)\|\dot{y}(t)+y(t)-y^*\|^2+\frac{1}{\rho\gamma}\|\bar{\lambda}(t)-\lambda^*\|^2 \\
&\quad\le c_9\|\dot{x}(t)+x(t)-x^*\|^2+c_{10}(\|\dot{x}(t)\|^2+\|\dot{y}(t)\|^2+\dot{\lambda}(t)\|^2),
\end{aligned}$$

where $c_9 = c_1\lambda_{\max}(Q)+(c_5+c_1c_6)(\rho\gamma)^{-1} > 0$, $c_{10} = c_2\lambda_{\max}(Q)+(c_7+c_8+c_2c_6)(\rho\gamma)^{-1} > 0$. This together with (5.5) and (5.6) implies that (5.7) holds with

$$\sigma = \max\left\{C_1(C_2+c_{10})^{-1}, \nu_f(\lambda_{\max}(\hat{P})+c_9)^{-1}\right\} > 0.$$

(iv). It follows from Lemma 5.3, (5.5) and (5.6) that (5.7) holds with

$$\sigma = \max\left\{C_1(C_2+\frac{c_7+c_8}{\rho\gamma})^{-1}, \nu_f(\lambda_{\max}(\hat{P})+\frac{c_5}{\rho\gamma})^{-1}, \nu_g(\lambda_{\max}(Q)+\frac{c_6}{\rho\gamma})^{-1}\right\} > 0.$$

$\square$

Table 1 summarizes the four cases under which we establish the exponential convergence of the dynamical system (3.1). In many practical problem models, the matrix B is usually equal to $-I_m$. In this case, B has a full column rank and $[B,A]$ has a full row rank. Here, we give an example to show that the conditions can be satisfied. Consider the elastic net problem:

$$\min_{x\in\mathbb{R}^n} \tau_1\|x\|_1+\frac{\tau_2}{2}\|x\|^2+\frac{1}{2}\|Mx-b\|^2,$$

where $\tau_1,\tau_2>0$. With the constraint $x=y$, the elastic net problem can be reformulated as:

$$\begin{aligned}
\min_{x,y}\quad & \tau_1\|y\|_1+\tfrac{\tau_2}{2}\|x\|^2+\tfrac{1}{2}\|Mx-b\|^2, \\
s.t.\quad & x-y=0,
\end{aligned}\tag{5.8}$$

It is easy to verify that condition (i) of Theorem 5.4 is satisfied. Then we can use the dynamical system (3.1) to solve (5.8) with exponential convergence.

It should be pointed out that the strong convexity of the objective function and the full rank assumption of the matrix may not be satisfied in some problems. For example, if we take $\tau_1 = 1$ and $\tau_2 = 0$, the elastic net problem becomes the following LASSO regression problem:

$$\begin{aligned}
\min_{x,y}\quad & \tau_1\|y\|_1+\tfrac{1}{2}\|Mx-b\|^2. \\
s.t.\quad & x-y=0,
\end{aligned}\tag{5.9}$$

where none of condition $(i)-(iv)$ is satisfied. This motivates us to consider the exponential converge of the dynamical system (3.1) without strong convexity assumption. Define the set-valued mapping $\mathscr{T}:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^p\rightrightarrows\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^p$, which is associated with the KKT system of the problem (1.1), as the following:

$$\mathscr{T}(x,y,\lambda) = \begin{pmatrix} A^T\lambda+\partial f(x) \\ B^T\lambda+\partial g(y) \\ Ax+By-b \end{pmatrix}.$$

It is easy to verify that

$$\Omega = \{(x,y,\lambda) | 0 \in \mathscr{T}(x,y,\lambda)\} = \mathscr{T}^{-1}(0). \tag{5.10}$$

Since Ω is a closed convex set, define

$$\mathrm{dist}(u(t),\Omega) = \min_{u\in\Omega} \|u(t)-u\| \quad \text{and} \quad \mathrm{dist}_G(u(t),\Omega) = \min_{u\in\Omega} \|u(t)-u\|_G,$$

where G is defined in Section 4. Then we have

$$\mathrm{dist}_G^2(u(t),\Omega) \le \lambda_{\max}(G)\mathrm{dist}^2(u(t),\Omega). \tag{5.11}$$

To investigate the exponential convergence without strong convexity, we make the following assumption.

Assumption 5.5. $\mathscr{T}$ is metrically subregular at (x^*,y^*,λ^*) for 0 with a neighbourhood U of (x^*,y^*,λ^*) and modulus $\kappa > 0$.

Remark that the metric subregularity condition is a standard assumption for the linear convergence of numerical algorithms for optimization problems when strong convexity assumption is not satisfied.

Theorem 5.6. *Suppose that Assumption 3.1, Assumption 3.2 and Assumption 5.5 hold. Then there exist $t_1 \ge t_0$ and $\sigma > 0$ such that*

$$dist_G^2(u(t),\Omega) \le e^{-\sigma(t-t_1)} dist_G^2(u(t_1),\Omega), \quad \forall t \ge t_1.$$

Proof. Consider the Lyapunov function

$$\mathscr{V}(t) = \mathrm{dist}_G^2(u(t),\Omega). \tag{5.12}$$

Denote $\hat{u}(t) = (\dot{x}(t)+x(t), \dot{y}(t)+y(t), \lambda(t))$. By Theorem 4.4, $\hat{u}(t) \to (x^*,y^*,\lambda^*) \in \Omega$ as $t \to +\infty$. Then there exists $t_1 \ge t_0$ such that $\hat{u}(t) \in U$ for all $t \ge t_1$. Since $\mathscr{T}$ is metrically subregular at (x^*,y^*,λ^*) for 0 with a neighbourhood U, it follows from (5.10) and the triangle inequality that for all $t \ge t_1$,

$$\begin{aligned} \mathrm{dist}^2(u(t),\Omega) &\le 2\mathrm{dist}^2(\hat{u}(t),\Omega) + 2\|\hat{u}(t)-u(t)\|^2 \\ &\le 2\kappa^2 \cdot \mathrm{dist}^2(0, \mathscr{T}(\dot{x}(t)+x(t), \dot{y}(t)+y(t), \lambda(t))) \\ &\quad + 2(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2). \end{aligned} \tag{5.13}$$

By the definition of $\mathscr{T}$, we can compute

$$\begin{aligned} &\mathrm{dist}^2(0, \mathscr{T}(\dot{x}(t)+x(t), \dot{y}(t)+y(t), \lambda(t))) \\ &\le \left\| \frac{1}{\gamma} A^T \dot{\lambda}(t) + P\dot{x}(t) \right\|^2 + \left\| \frac{1}{\gamma} B^T \dot{\lambda}(t) + Q\dot{y}(t) - \rho B^T A \dot{x}(t) \right\|^2 + \left\| \frac{1}{\rho\gamma} \dot{\lambda}(t) \right\|^2 \\ &\le C_3(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2 + \|\dot{\lambda}(t)\|^2), \end{aligned} \tag{5.14}$$

where the first inequality follows from (3.1), (4.8) and (4.9),

$$C_3 = \max\left\{ 2\|P\|^2 + 3\|\rho B^T A\|^2, 3\|Q\|^2, \frac{1}{\gamma^2}(2\|A\|^2 + 3\|B\|^2 + \frac{1}{\rho^2}) \right\} > 0.$$

It follows from (5.11)-(5.14) that for all $t \ge t_1$,

$$\mathscr{V}(t) \le C_4(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2 + \|\dot{\lambda}(t)\|^2), \tag{5.15}$$

where $C_4 = (2\kappa^2 C_3 + 2)\lambda_{\max}(G) > 0$. Since $G \succ 0$, there exists an invertible matrix $\bar{G}$ such that $G = \bar{G}^T \bar{G}$. Then we have

$$\mathscr{V}(t) = \min_{u\in\Omega} \|u(t) - u\|_G^2 = \min_{u\in\Omega} \|\bar{G}u(t) - \bar{G}u\|^2 = \operatorname{dist}^2(\bar{G}u(t), \bar{\Omega}),$$

where $\bar{\Omega} = \{\bar{G}u \mid u \in \Omega\}$ is a nonempty closed convex set. By [28, Corollary 12.30], we obtain

$$\begin{aligned}\dot{\mathscr{V}}(t) &= \frac{d(\operatorname{dist}^2(\bar{G}u(t), \bar{\Omega}))}{dt} \\ &= 2\langle \bar{G}u(t) - P_{\bar{\Omega}}(\bar{G}u(t)), \bar{G}\dot{u}(t)\rangle \\ &= 2\langle u(t) - \bar{G}^{-1}P_{\bar{\Omega}}(\bar{G}u(t)), G\dot{u}(t)\rangle,\end{aligned}$$

where

$$P_{\bar{\Omega}}(u) = \arg\min_{\bar{u}\in\bar{\Omega}} \|u - \bar{u}\|$$

is the projection operator onto $\bar{\Omega}$. Since $\bar{G}^{-1}P_{\bar{\Omega}}(\bar{G}u(t)) \in \Omega$, replacing u^* by $\bar{G}^{-1}P_{\bar{\Omega}}(\bar{G}u(t))$ in (4.7), from (5.5) we have

$$\dot{\mathscr{V}}(t) \le -2C_1(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2 + \|\dot{\lambda}(t)\|^2), \quad \forall t \ge t_0. \tag{5.16}$$

It follows from (5.15)-(5.16) that for all $t \ge t_1$,

$$\dot{\mathscr{V}}(t) + \sigma V(t) \le 0$$

with $\sigma = 2C_1/C_4$. Then by similar arguments as in Theorem 5.4, we obtain the desired result. □

Remark 5.7. When f and g are both piecewise linear-quadratic functions, $\mathscr{T}$ is a piecewise linear multifunction. The desired metric subregularity of $\mathscr{T}$ (Assumption 5.5) follows immediately from [35, Theorem 3.3]. In particular, Assumption 5.5 holds for the LASSO regression problem (5.9).

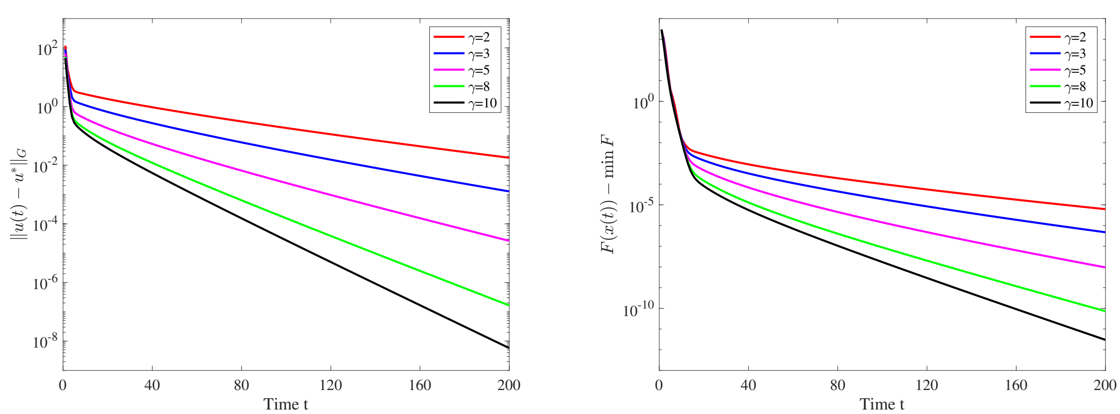

FIGURE 1. Simulation of dynamical system (3.1) for the elastic net problem with different γ and $P = (\frac{1}{2\gamma} - 1)I_n$

6. Numerical stimulation

In this section, two examples are simulated to verify and visualize the theoretical results established in Section 5. We always use Matlab ODE solver *ode*45 to solve dynamical systems.

Consider the elastic net problem (5.8) with $\tau_1 = 1$ and $\tau_2 = 0.1$. Let $M \in \mathbb{R}^{40\times 100}$ be generated by the standard Gaussian distribution. Suppose that $x^* \in \mathbb{R}^{100}$ has 5 non-zero elements generated randomly and $b = Mx^*$. Now, we using the dynamical system (3.1) to solve the problem (5.8). Set the parameter $Q = 0.2I_m$ and $\rho = 1$. Figure 1 shows the behavior of $\|u(t) - u^*\|_G$ and $F(x(t)) - \min F$ with different choice of γ and $P = (\frac{1}{2\gamma} - 1)I_n$, where $F(x) = \tau_1\|x\|_1 + \frac{\tau_2}{2}\|x\|^2 + \frac{1}{2}\|Mx - b\|^2$.

Consider the LASSO regression problem (5.9) with $\tau_1 = 1$. Let $M \in \mathbb{R}^{100\times 200}$ be generated by the standard Gaussian distribution. Suppose that $x^* \in \mathbb{R}^{200}$ has 10 non-zero elements generated randomly and $b = Mx^*$. Using the dynamical system (3.1) to solve the problem (5.9) with the parameter $Q = 0.2I_m$ and $\rho = 1$, $P = -0.75I_n$. Figure 2 shows the behavior of $\|u(t) - u^*\|_G$ and $F(x(t)) - \min F$ with different choice of γ, where $F(x) = \|x\|_1 + \frac{1}{2}\|Mx - b\|^2$.

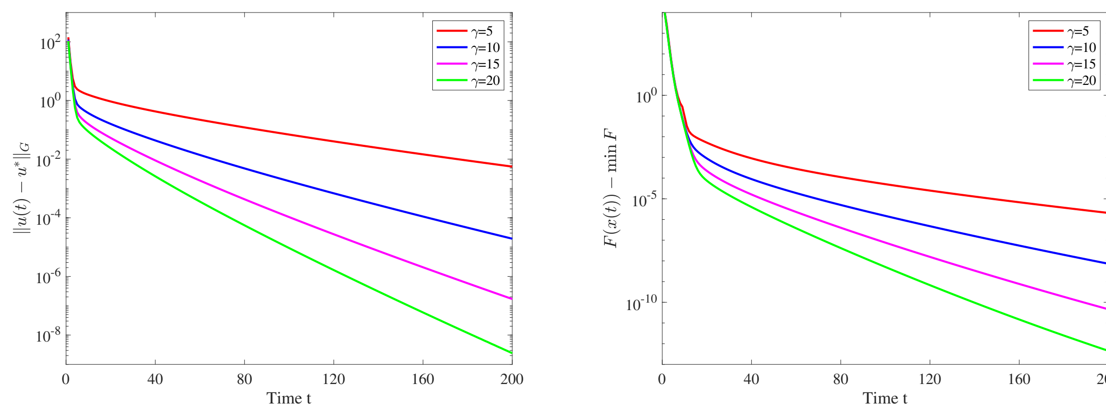

FIGURE 2. Simulation of dynamical system (3.1) for the LASSO regression problem with different γ

From Figure 1 and Figure 2, we can see that $\|u(t) - u^*\|_G$ and $F(x(t)) - \min F$ converge to zero approximately with a linear rate, which matches our theoretical analysis.

7. Conclusion

In this paper, we propose a new primal-dual dynamical system approach (3.1) for solving the separable convex optimization problem, which can be treated as the continuous limit of the discrete proximal ADMM algorithm with a larger dual step size. The trajectory of the proposed dynamical system globally converges to a saddle point of the problem under suitable parameter settings. The decay rate of the trajectory can be exponential provided that additional strong convexity and full rank assumptions are satisfied. The exponential convergence of the trajectory to the saddle point set can be guaranteed under the metric subregularity condition, instead of strong convexity and full rank assumptions. Two numerical results support our theoretical results.

Acknowledgments
The authors would like to thank the referees and the editor for their helpful comments and suggestions on earlier versions of this paper, which have led to the improvement of the current paper.

REFERENCES

[1] A. Beck, First-order Methods in Optimization, SIAM, 2017.
[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2010)1-122.
[3] D. Gabay, Augmented Lagrangian methods: applications to the numerical solution of boundary-value problemsm volume 15 of Studies in Mathematics and Its Applications, Elsevier, 1983.
[4] Z. Lin, H. Li, C. Fang, Accelerated Optimization for Machine Learning, Springer, Singapore, 2019.
[5] M. Raus, Y. Elshiaty, S. Petra, Accelerated Bregman divergence optimization with SMART: An information geometric point of view, J. Appl. Numer. Optim. 6 (2024) 1-40.
[6] T. Goldstein, B. O'Donoghue, S. Setzer, R. Baraniuk, Fast alternating direction optimization methods, SIAM J. Imaging Sci.7(2014)1588-1623.
[7] L. Tan, K. Guo, Bregman ADMM: A new algorithm for nonconvex optimization with linear constraints, J. Nonlinear Var. Anal. 9 (2025), 179-196.
[8] F. Battahi, Z. Chbani, H. Riahi, On the simultaneous convergence of values and trajectories of continuous inertial dynamics with Tikhonov regularization to solve convex minimization with affine constraints, Appl. Set-Valued Anal. Optim. 7 (2025) 373-392.
[9] X. He, R. Hu, Y.P. Fang, Convergence rates of inertial primal-dual dynamical methods for separable convex optimization problems, SIAM J. Control Optim. 59(2021) 3278-3301.
[10] X. He, R. Hu, Y.P. Fang, "Second-order primal" + "first-order dual" dynamical systems with time scaling for linear equality constrained convex optimization problems, IEEE Trans. Automat. Control. 67 (2022) 4377-4383.
[11] X. He, R. Hu, Y.P. Fang, Fast primal-dual algorithm via dynamical system for a linearly constrained convex optimization problem, Automatica 146 (2022) 110547.
[12] X. He, R. Hu, Y.P. Fang, Inertial primal-dual dynamics with damping and scaling for linearly constrained convex optimization problems, Appl. Anal. 102 (2023) 4114-4139.
[13] D. Feijer, F. Paganini, Stability of primal–dual gradient dynamics and applications to network optimization, Automatica 46 (2010) 1974-1981.
[14] S. Liang, G. Yin, Exponential convergence of distributed primal–dual convex optimization algorithm without strong convexity, Automatica 105 (2019) 298-306.
[15] H. Luo, A primal-dual flow for affine constrained convex optimization, 2021; arXiv:2103.06636.
[16] X. Zeng, P. Yi, Y. Hong, Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach, IEEE Trans. Automat. Control. 62 (2016) 5227-5233.
[17] Y. Zhu, W. Yu, G. Wen, G. Chen, Projected primal–dual dynamics for distributed constrained nonsmooth convex optimization, IEEE Trans. Cybern. 50 (2018) 1776-1782.
[18] G. Qu, N. Li, On the exponential stability of primal-dual gradient dynamics, IEEE Contr. Syst. Lett. 3 (2018) 43-48.
[19] Y. Tang, G. Qu, N.Li, Semi-global exponential stability of augmented primal–dual gradient dynamics for constrained convex optimization, Systems Control Lett. 144 (2020) 104754.
[20] G. França, D.P. Robinson, R. Vidal, ADMM and accelerated ADMM as continuous dynamical systems, In International Conference on Machine Learning. pp. 1559-1567, 2018.
[21] R.I. Boţ, E.R. Csetnek, S. László, A primal-dual dynamical approach to structured convex minimization problems, J. Differential Equations, 269 (2020) 10717-10757.
[22] S. Bitterlich, E.R. Csetnek, G. Wanka, A dynamical approach to two-block separable convex optimization problems with linear constraints, Numer. Funct. Anal. Optim. 42 (2020) 1-38.
[23] M. Chao, L. Liu, A dynamical alternating direction multiplier method for two-block optimization problems, Nonlinear dynamics 111 (2023) 6557–6583.

[24] D. Ding, M.R. Jovanović, Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented Lagrangian, In: 2019 American Control Conference (ACC), pp. 3414-3419, 2019.
[25] S. Hassan-Moghaddam, M.R. Jovanović, Proximal gradient flow and Douglas–Rachford splitting dynamics: global exponential stability via integral quadratic constraints, Automatica 123 (2021) 109311.
[26] X. Chen, M. Li, Exponential stability of primal-dual gradient dynamics with non-strong convexity, In: 2020 American Control Conference (ACC), pp. 1612-1618, 2020.
[27] Z. Wang, W. Wei, C. Zhao, Z. Ma, Z. Zheng, Y. Zhang, F. Liu, Exponential stability of partial primal–dual gradient dynamics with nonsmooth objective functions, Automatica 129 (2021) 109585.
[28] H.H. Bauschke, P.L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces: Second Edition, Springer, New York, 2016.
[29] H. Brezis, Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holl. Math. Stud. vol. 5, North-Holland/Elsevier, New York, 1973.
[30] L. Dontchev, R.T.Rockafellar, Implicit Functions and Solution Mappings, Springer, 2009.
[31] B. Abbas, H. Attouch, B.F. Svaiter, Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces, J. Optim. Theory Appl. 161 (2014) 331-360.
[32] W. Deng, W. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, J. Sci. Comput. 66 (2016) 889-916.
[33] X. Yuan, S. Zeng, J. Zhang, Discerning the linear convergence of ADMM for structured convex optimization through the lens of variational analysis, J. Mach. Learn. Res. 21 (2020) 1-75.
[34] M.N. Dao, H.M.Phan, Adaptive Douglas-Rachford splitting algorithm for the sum of two operators, SIAM J. Optim. 29 (2019) 2697-2724.
[35] X.Y. Zheng, K.F. Ng, Metric subregularity of piecewise linear multifunctions and applications to piecewise linear multiobjective optimization, SIAM J. Optim. 24 (2014) 154-174.