

A PROJECTION-FREE DYNAMICS FOR NONSMOOTH COMPOSITE OPTIMIZATION

WEI NI, YANGFAN QIU, AND YANYAN XIAO

This paper proposes a projection-free primal-dual dynamics for the nonsmooth composite optimization problems with equality and inequality constraints. To deal with optimization constraints, this paper departs from the use of gradient projection method, but resorts to the idea of mirror descent to design a continuous-time smooth optimization dynamics which advantageously leads to easier convergence analysis and more efficient numerical simulation. Also, the strategy of proximal augmented Lagrangian (PAL) is extended to incorporate general convex equality-inequality constraints and the strong convexity-concavity of the primal-dual variables is achieved, ensuring exponential convergence of the resulting algorithm. Furthermore, the convergence result in this paper extends existing exponential convergence which either takes no account of constraints or considers only affine linear constraints, and it also enhances existing asymptotic convergence under convex constraints which unfortunately depends on the complex gradient projection scheme.

Keywords: composite optimization, proximal augmented Lagrangian, projection-free

Classification: 90C25, 90C30, 90C90

1. INTRODUCTION

Composite optimization with nonsmooth regularization has been a topic of a large amount of research. This problem consists in minimizing the sum of a differentiable function $f(x)$ and a possibly nonsmooth one $\varphi(Tx)$, with the latter being a composition of an extended-valued function φ with a linear operator T . A prominent example is the empirical risk minimization with nonsmooth regularization as parameter choice rule in deep neural network [44]. As another example, the typical signal processing problem of restoration of a continuous signal from its digital counterparts is tackled by solving an optimization problem known as the lasso [65] which is a specific case of the problem considered in this paper. Applications in other fields include inverse problem [7], statistical estimation [10], and control [46]. Due to the presence of the nondifferentiable part φ in the objective function, classical gradient based optimization algorithms, such as gradient descent, conjugate gradient, and Newton methods, cannot be directly applied. To tackle non-smoothness, a direct research line is the nonsmooth optimization method, including subgradient method [62], stochastic approximation [50] and bundle method

[47], but the convergence is very slow. As a parallel, the smooth optimization method [52] replaces the nonsmooth component with smooth approximations parameterized by a positive parameter and then resorts to smooth optimization routines; however, as the parameter approaches zero, the problem becomes increasingly involved. A breakthrough was made via proximal operator in a series of milestone papers which do not require explicit gradient [7, 14, 60, 66]. In these methods, the smooth part of the objective function is processed by evaluation of its gradient operator, while a nonsmooth part is handled via its Moreau proximal operator [49], resulting in iterations of the form $x^+ \in \text{prox}_{\gamma\psi(T(x))}(x - \gamma\nabla f(x))$ for some suitable stepsize γ . Proximal method can be derived from operator splitting techniques (see Douglas [27], Passty [55] and Bruck [17]) which find their roots in the works of Browder [16] and Minty [48], and now it is developed in the more general setting of maximal monotone operators [1]. The more recent work by Combettes and Wajs [23] provides important insights into the proximal method and has popularized the method for a wide audience. Nowadays the proximal method becomes an important part of convex analysis, optimization, and nonlinear functional analysis.

The above proximal method involves the proximal calculation with respect to the composite function $\varphi \circ T$ which is however difficult to evaluate for general case. To overcome this, a frequently used technique is the variable splitting which constructs a new variable y to serve as the argument of φ and imposes additional constraint $y = Tx$. Then this constrained optimization problem can be tackled by using an augmented Lagrangian (AL[†]) method [54] originally due to Hestenes [39] and Powell [56] and by designing corresponding primal-dual algorithms [4]. The AL here extends the classical Lagrangian by including an extra quadratic term, without changing the optimal solution but facilitating the proximal operator to be used here. The proximal version of this method is earlier developed by Rockafellar [60] and its connection to alternating direction of multiplier method is recently explored by Boyd [14], followed by a burst of research activities in subsequent years. Note that the AL function mentioned above still contains a nondifferentiable term and thus precludes the standard strategy of gradient descent on primal variables and gradient ascent on dual variables. To conquer the non-differentiability in the traditional AL, Dhingra et al. [24] recently propose a new kind of differentiable Lagrangian, called PAL, which is obtained by taking infimum over the AL with respect to the primal variable of the nonsmooth function. Compared with the AL in [54], a coming advantage of the PAL is that it is continuously differentiable on both primal and dual variables so that direct primal-dual dynamics can be applied on the PAL.

With the help of PAL, the original optimization problem is transformed into the computation of the saddle point of this PAL, and various formulations of primal-dual dynamics have been explored in the literature with focus on investigating asymptotic or exponential convergence under different conditions [25, 28, 32, 33, 37, 41, 64]. In a broader view, ever since von Neumann's celebrated min-max theorem [53], the problem of

[†]For convenience, other acronyms defined in this paper are also summarized here:

AL: augmented Lagrangian,	PAL: proximal augmented Lagrangian,
IQC: integral quadratic constraints,	SCQ: Slater's constrained qualification,
KKT: Karush-Kuhn-Tucker,	LICQ: linear independence constraint qualification.

finding a saddle point for a convex-concave function has been a major focus of disciplines like machine learning [34], robust optimization [8], robust control [38] and game theory [6]. For the abstract gradient descent-ascent algorithm which performs gradient descent in convex variable and gradient ascent in concave variable, there is a vast literature obtaining asymptotic convergence if either the convexity or concavity are strict [4, 19, 43, 63]; exponential convergence can be further ensured if strongly convex strongly concave condition is imposed [19]. However, when specialized to the case of classical Lagrangian which is only linear on the dual variables, exponential stability can not be obtained directly for the primal-dual dynamics. This has also been highlighted in [68] that the combination of general convex constraints and nonsmooth objectives poses a significant challenge to achieving exponential stability. To prove exponential stability under extra conditions, other routines include the contraction analysis [22] and the technique of spectral bounds of saddle matrices [9]. In this paper, instead of building on the classical Lagrangian but on the PAL, we develop a primal-dual optimization dynamics. Aside from the advantage of variable splitting inherent in the PAL, we will show that the PAL induces strong concavity on dual variable under appropriate conditions and renders the corresponding primal-dual optimization dynamics to achieve exponential stability by taking the method of Lyapunov function and convex analysis. Obviously, improved upon asymptotic stability, exponential stability offers explicit convergence rate of the optimization algorithms.

When considering constraints in the PAL method, most existing results have been successful in proving exponential stability either for equality constraints [22, 68] or for (affine) linear inequality constraints [2, 57]. When only asymptotic stability is required, general convex inequality constraints can be tackled either by making use of the framework of hybrid dynamical system [30] or by referring to techniques involving projection [33, 64], where the former invokes a generalized Lasalle invariance principle developed for hybrid systems and the latter utilizes the sophisticated nonsmooth analysis. This generalized Lasalle invariance principle is also improved to be used for the projection approach; see Cherukuri et al. [20, 21]. Unfortunately, both the hybrid system approach and the projection approach would result in discontinuity of the primal-dual dynamics and could bring difficulties in the convergence proof. To overcome this difficulty, this paper pursues a projection free method to design a smooth primal-dual optimization dynamics by utilizing the technique of mirror descent [51, 58] under the general framework of barrier operator [12]. Our method avoids projection and thus reduces the difficulties of convergence analysis and iterative computation.

Aside from merit of yielding a smooth optimization dynamics, further advantage of our method lies in its ability to deal with general nonlinear convex optimization constraints. For optimization algorithm design and corresponding exponential convergence analysis, a popular choice is the IQC framework [29, 45] which provides linear matrix inequality based conditions for exponential stability of the feedback interconnection of a linear system with a nonlinearity satisfying IQC. Although the gradient mapping or the proximal mapping individually satisfies IQC, composition of each with a nonlinear map generally does not satisfy IQC; instead, IQC holds for compositions with affine operations [29]. In view of this fact, most exponential convergence of PAL methods based on IQC deals with either equality constraints [24, 36, 37, 68, 26] (which should

be affine in decision variables in view of the convex nature of the problem) or affine inequality constraints [5, 35], with general nonlinear convex inequality constraints not touched. Similar situations occur in other literature; for one example in [2], the authors proposed another projection-free method to deal with inequality constraints, but only limited to affine inequality constraints and for another in [57], exponential stability is obtained but limited only to affine inequality constraints. Our method does not rely on IQC analysis and exponential stability can be achieved under general convex inequality constraints. Although general convex inequality constraints are considered in [33] and [11, 64], the former can only guarantee asymptotic convergence and the latter attains semi-globally exponential convergence. These deficiencies of asymptotic convergence and semi-global exponential convergence are improved to exponential convergence in this paper. We also note the works [2, 11, 57] which handle inequality constraints without explicit projection. However, [2, 11, 57] replace projection-based updates with max-based regularization (specifically algorithm (9a)–(9c) in [2], (9a)–(9b) in [57], and (1b) in [11]), implicitly ensuring nonnegativity of the multiplier associated with inequality constraint. While this improves computational simplicity, the max operator still makes the optimization dynamics nonsmooth in the classical sense. In contrast, our projection-free dynamics is fully smooth, and this smooth nature enables rigorous stability certification via classical Lyapunov techniques, bypassing the need for nonsmooth analysis tools like differential inclusions. More important, the max-based method in [2, 57] are limited to affine inequality constraints, with no extension to general convex inequality constraints. Our approach overcomes this limitation and extends to general convex inequality constraints.

To summarize, this paper proposes a projection-free method for nonsmooth composite optimization with constraints. The contributions of this paper are as follows.

- Firstly, equality and general inequality optimization constraints are taken into the nonsmooth composite optimization problem and a projection-free optimization dynamics is proposed. Unlike existing projection or max based methods of dealing with constraints which render nonsmooth optimization dynamics, the optimization dynamics proposed in this paper is smooth, with the benefit of reducing the difficulties of convergence analysis and iteration complexity, bypassing the need for nonsmooth analysis tools like differential inclusions.
- Secondly, improved on the classical Lagrangian which is only linear on the dual variable, the PAL method is adopted and it extends existing PAL method by considering optimization constraints. Strong concavity on dual variable is obtained in this paper so that exponential stability is achieved, thereby improving upon existing results that offer only asymptotic stability. In contrast to asymptotic stability, exponential stability provides explicit convergence rates for the optimization algorithms.
- Lastly, the method of achieving exponential stability in this paper differs existing IQC-based approaches that work only for equality constraints or affine inequality constraints. our method does not rely on IQC analysis and is capable of handling general nonlinear convex inequality constraints while still ensuring exponential convergence, going beyond the limitations of existing IQC-based approaches.

The rest of this paper is structured as follows. Preliminaries are put in Section 2.

Section 3 describes the nonsmooth and composite optimization problem. Section 4 is devoted the design of the optimization dynamics and the analysis of algorithm convergence. A simulation example is introduced in Section 5 and a brief conclusion is made in Section 6.

2. PRELIMINARIES

We use $\mathbb{R}^n$ to denote the set of n -dimensional real-valued vectors and $\mathbb{R}_+^n$ those vectors with nonnegative components. For $a \in \mathbb{R}^n$, by $a \prec 0$ ($a \preceq 0$) we mean that each entry of a is less than (less than or equal to) zero. Euclidian norm of a vector is denoted by $\|\cdot\|$. Letting $a = (a_1, \dots, a_n)^\top \in \mathbb{R}^n$ and $b = (b_1, \dots, b_n)^\top \in \mathbb{R}^n$, we define $a \odot b = (a_1 b_1, \dots, a_n b_n)^\top$ and $a \oslash b = (a_1/b_1, \dots, a_n/b_n)^\top$. For vectors $\alpha_1, \dots, \alpha_m$, the notation $\text{col}\{\alpha_i\}_{i=1}^m$ denotes a new vector $(\alpha_1^\top, \dots, \alpha_m^\top)^\top$.

An extended value function $\varphi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ with effective domain $\text{dom}(\varphi) \equiv \{y \mid \varphi(y) < \infty\}$ is proper if $\varphi(y) < +\infty$ for at least one $y \in \text{dom}(\varphi)$ and $\varphi(y) > -\infty$ for all $y \in \text{dom}(\varphi)$. A function $\varphi : \mathbb{R}^m \rightarrow [-\infty, \infty]$ is lower semicontinuous if $\{\varphi \leq c\} \triangleq \{y \in \mathbb{R}^m : \varphi(y) \leq c\}$ is a closed subset of $\mathbb{R}^m$ for all $c \in \mathbb{R}$. A function $\varphi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ is closed if its epigraph $\{(y, c) \in \mathbb{R}^m \times \mathbb{R} \mid \varphi(y) \leq c\}$ is a closed set. For a convex and lower semicontinuous function φ , its conjugate function (or Fenchel-Legendre transform) $\varphi^* : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined as $\varphi^*(v) = \sup_{u \in \text{dom}(\varphi)} \{uv - \varphi(u)\}$, which is also convex and lower-semicontinuous. Clearly only those u in $\text{dom}(\varphi)$ are relevant in the calculation of this supremum.

For an extended value function $\varphi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$, its subdifferential at the point $y \in \text{dom}(\varphi)$ is denoted as $\partial\varphi(y) = \{v \in \mathbb{R}^y \mid \varphi(z) - \varphi(y) \geq \langle v, z - y \rangle \text{ for all } z\}$ which is always closed and convex for $y \in \text{int}(\text{dom } \varphi)$; if φ is convex then it is nonempty and if φ is differentiable then $\partial\varphi(y) = \{\nabla\varphi(y)\}$, the usual gradient.

A differentiable function f is m_f -strongly convex if for any x and $\hat{x}$, $f(\hat{x}) \geq f(x) + \langle \nabla f(x), \hat{x} - x \rangle + \frac{m_f}{2} \|\hat{x} - x\|_2^2$, or equivalently, $\|\nabla f(\hat{x}) - \nabla f(x)\|_2 \geq m_f \|\hat{x} - x\|_2$, or equivalently, $x \rightarrow f(x) - \frac{m_f}{2} \|x\|_2^2$ is convex. The gradient of a continuously-differentiable function f is Lipschitz continuous with parameter ℓ_f if for any x and $\hat{x}$, $f(\hat{x}) \leq f(x) + \langle \nabla f(x), \hat{x} - x \rangle + \frac{\ell_f}{2} \|\hat{x} - x\|_2^2$, or equivalently, $\|\nabla f(x) - \nabla f(\hat{x})\|_2 \leq \ell_f \|x - \hat{x}\|_2$. Moreover, if an m_f -strongly convex function f has an ℓ_f -Lipschitz continuous gradient, then any x and $\hat{x}$ it holds $\langle \nabla f(x) - \nabla f(\hat{x}), x - \hat{x} \rangle \geq \frac{m_f \ell_f}{m_f + \ell_f} \|x - \hat{x}\|_2^2 + \frac{1}{m_f + \ell_f} \|\nabla f(x) - \nabla f(\hat{x})\|_2^2$. We say that f is strongly concave if $-f$ is strongly convex.

Let $K \subseteq \mathbb{R}^{n+m}$ be nonempty, closed and convex set. We say that a function $\varphi(x, y) : K \rightarrow \mathbb{R}$ is concave-convex on K if for any $(x, y) \in K$, $\varphi(x, y)$ is a concave function of y and $\varphi(x, y)$ is a convex function of x . If either the concavity or convexity is always strict, we say that is strictly concave-convex on K . For a concave-convex function $\varphi : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, we say that $(\bar{x}, \bar{y}) \in \mathbb{R}^{n+m}$ is a saddle point if for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ we have the inequality $\varphi(x, \bar{y}) \leq \varphi(\bar{x}, \bar{y}) \leq \varphi(\bar{x}, y)$. If φ is in addition continuously differentiable, then $(\bar{x}, \bar{y})$ is a saddle point if and only if $\nabla_{\bar{x}} \varphi(\bar{x}, \bar{y}) = 0$ and $\nabla_{\bar{y}} \varphi(\bar{x}, \bar{y}) = 0$.

For a non-differentiable extended value function $\varphi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$, its Moreau envelope φ_μ is defined as

$$\varphi_\mu(v) \triangleq \min_{y \in \mathbb{R}^m} \left\{ \varphi(y) + \frac{1}{2\mu} \|y - v\|^2 \right\} \leq \varphi(v). \quad (2.1)$$

Even though φ may not be differentiable, its Moreau envelope may be, with $\text{dom}(\varphi_\mu) = \mathbb{R}^m$. We also note that φ_μ is convex since the function in the curly brackets is jointly convex in (x, y) and thus the epigraph of φ_μ is the projection of a convex set. The Moreau envelope φ_μ provides a smooth approximation of φ from below and retains the same minimizers [61]. The minimizer of the problem (2.1) is called as the proximal operator which is defined and calculated as follows,

$$\text{Prox}_{\mu\varphi}(v) = \underset{y \in \mathbb{R}^m}{\text{argmin}} \left\{ \varphi(y) + \frac{1}{2\mu} \|y - v\|^2 \right\} \quad (2.2)$$

$$= y - \mu \nabla \varphi_\mu(y), \quad (2.3)$$

where the second equality can be seen by noting that the gradient of the Moreau is

$$\nabla \varphi_\mu(y) = \frac{1}{\mu} [y - \text{Prox}_{\mu\varphi}(v)].$$

The proximal operator expressed in the equality (2.3) can be viewed as a gradient descent for the Moreau envelope φ_μ . It can also be viewed as a generalization of the notion of Euclidean projection if one takes φ to be the indicator function of some convex set C so that $\text{Prox}_{\mu\varphi}(y) = \underset{z \in C}{\text{argmin}} \|y - z\|^2$. For some choices of φ , the corresponding $\text{Prox}_{\mu\varphi}$ has well known closed forms. For example, if $\varphi(x) = \|x\|_1$, then $\text{Prox}_{\mu\varphi} = \text{soft}(y, \mu)$, where $\text{soft}(\cdot, \mu)$ denotes the component-wise application of the soft-threshold function $y \mapsto \text{sign}(y) \max\{|y| - \mu, 0\}$. In general the proximal operator may be set-valued, but it is scalar-valued if $f(x)$ is proper and convex.

3. PROBLEM FORMULATION

Consider the following nonsmooth composite optimization problem with equality and inequality constraints:

$$\mathcal{P} : \begin{cases} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) + \varphi(Tx), \\ \text{subject to} & g(x) \preceq 0, \\ & h(x) = 0, \end{cases} \quad (3.1a)$$

$$g(x) \preceq 0, \quad (3.1b)$$

$$h(x) = 0, \quad (3.1c)$$

where $x \in \mathbb{R}^n$ is the decision variable, $T \in \mathbb{R}^{m \times n}$ is a given matrix of full column rank, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\varphi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, convex, and lower semicontinuous, with f continuously differentiable and φ non-differentiable, $g = (g_1, \dots, g_r)^\top : \mathbb{R}^n \rightarrow \mathbb{R}^r$, and $h = (h_1, \dots, h_s)^\top : \mathbb{R}^n \rightarrow \mathbb{R}^s$ are respectively the inequality and the equality constraints, with $g_j, j = 1, \dots, r$ being convex and continuously differentiable and $h_j, j = 1, \dots, s$ being affine. This makes the optimization problem (3.1) a convex one. In the problem (3.1), the T can represent a convolution operator for the image deconvolution problem and represents operation of tomographic projections in the general image reconstruction problems [3].

By introducing an auxiliary variable $y \triangleq Tx$ and by imposing additional constraint $y = Tx$, the original optimization problem (3.1) can be equivalently transformed into

the following one

$$\tilde{\mathcal{P}} : \begin{cases} \text{minimize} & f(x) + \varphi(y), \\ \text{subject to} & g(x) \preceq 0, \\ & h(x) = 0, \\ & Tx = y. \end{cases} \quad \begin{aligned} (3.2a) \\ (3.2b) \\ (3.2c) \\ (3.2d) \end{aligned}$$

Clearly, (x^*, y^*) is an optimal solution to (3.2) if and only if x^* is an optimal solution to the original optimization (3.1). The above procedure of creating a new variable y to serve as the argument of φ and imposing the constraint $y = Tx$ is a standard technique called variable splitting. As will be seen, the transformed optimization problem (3.2) is easier to solve than the original one (3.1). The following assumptions are made throughout this paper.

Assumption 1. The function f is α -strongly convex for some $\alpha > 0$ and continuously differentiable.

Assumption 2. The function φ is a convex and subdifferentiable function whose subgradient is $1/\ell$ -Lipschitz continuous for some $\ell > 0$; that is, for all $y, \tilde{y} \in \mathbb{R}^m$, it has $[d(y) - d(\tilde{y})]^T (y - \tilde{y}) \leq 1/\ell \|y - \tilde{y}\|^2$, where $d(y) \in \partial g(y)$ and $d(\tilde{y}) \in \partial g(\tilde{y})$.

Let (x^*, y^*) be an optimal solution to the problem (3.2). By assuming additional assumptions on the constraint functions, called constrained qualifications, the following Karush–Kuhn–Tucker (KKT) systems hold at the minimizer (x^*, y^*) : there exist multipliers $(\lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$ such that

$$\begin{aligned} KKT : \quad & \begin{cases} g(x^*) \preceq 0, & (3.3a) \\ h(x^*) = 0, & (3.3b) \\ Tx^* = y^*, & (3.3c) \\ \lambda^* \succeq 0, & (3.3d) \\ \lambda^* \odot g(x^*) = 0, & (3.3e) \\ \nabla f(x^*) + \partial \phi(y^*) + \sum_{i=1}^r \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^s \lambda_i^* \nabla h_i(x^*) + T^\top \tilde{\lambda}^* \ni 0. & (3.3f) \end{cases} \end{aligned}$$

A widely used constrained qualification is the Slater’s constrained qualification (SCQ): there exist $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $g(x) \prec 0$, $h(x) = 0$ and $Tx = y$. In other words, assuming SCQ, “ (x^*, y^*) solves $(\tilde{\mathcal{P}})$ ” $\Rightarrow$ “ $\exists$ a set of $(\lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*)$ together with (x^*, y^*) solving (KKT) ”. Furthermore, for convex problem, this implication is bidirectional. Refer to [13] for details. Although SCQ guarantees existence of multipliers satisfying the KKT system (3.3), it does not ensure uniqueness. A stronger constrained qualification, named the linear independence constraint qualification (LICQ), can achieve this goal. Using $\nabla_J g(x^*)$ to denote the submatrix of $\nabla g(x^*)$ given by rows with indices in $J(x^*) = \{i | g_i(x^*) = 0\}$, the LICQ is defined as

$$\text{rank} \begin{bmatrix} \nabla h(x^*) \\ \nabla_J g(x^*) \end{bmatrix} = s + |J(x^*)|, \quad (3.4)$$

here $|\cdot|$ denotes the set cardinality. By assuming LICQ, one obtains a result [67] on the existence and uniqueness of multipliers satisfying (3.3),

$$“(x^*, y^*) \text{ solves } (\tilde{P})” \Rightarrow “\exists! (\lambda_{ij}^*, \lambda_{ij}^*, \lambda_{ij}^*) \text{ together with } (x^*, y^*) \text{ solving (KKT)}”, \quad (3.5)$$

where the notation $\exists!$ stands for “there exists a unique”. This direction is bidirectional if the problem is convex. In view of above considerations, the following assumption is made.

Assumption 3. The constraint functions in (3.2b)–(3.2c) satisfy LICQ.

4. OPTIMIZATION DYNAMICS AND CONVERGENCE ANALYSIS

This section proposes an exponentially stable primal-dual dynamical system without using gradient projection to solve the optimization problem (3.1) or (3.2) building on the construction of a PAL.

4.1. Design of PAL

For the optimization problem (3.2), define the following AL function

$$L^\mu(x, y; \lambda, \lambda, \lambda) = f(x) + \varphi(y) + \lambda^\top g(x) + \lambda^\top h(x) + \lambda^\top (Tx - y) + \frac{1}{2\mu} \|Tx - y\|^2, \quad (4.1)$$

where $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ are the primal variables and $(\lambda, \lambda, \lambda) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$ are the dual variables. Under the LICQ conditions in Assumption 3, (x^*, y^*) is an optimal solution to the optimization problem (3.2) if and only if there exists $(\lambda^*, \lambda^*, \lambda^*) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$ such that the following saddle point inequality

$$L^\mu(x^*, y^*; \lambda, \lambda, \lambda) \leq L^\mu(x^*, y^*; \lambda^*, \lambda^*, \lambda^*) \leq L^\mu(x, y; \lambda^*, \lambda^*, \lambda^*) \quad (4.2)$$

holds for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $(\lambda, \lambda, \lambda) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$.

The saddle point property (4.2) leads us naturally to consider designing naive primal-dual dynamics based on the augmented Lagrangian L^μ with gradient descent on (x, y) and gradient ascent on $(\lambda, \lambda, \lambda)$, with the aim to seek the optimal solution (x^*, y^*) of the optimal problem (3.2). However, this naive primal-dual dynamics can only achieve asymptotic stability since L^μ is only linear on $(\lambda, \lambda, \lambda)$; to achieve exponential stability of primal-dual dynamics, strong concavity on $(\lambda, \lambda, \lambda)$ should be required. Toward this goal, the subsequent paragraph introduces a PAL which is strongly concave in $(\lambda, \lambda, \lambda)$.

Note that the AL in (4.1) can be rewritten via completing of squares as

$$L^\mu(x, y; \lambda, \lambda, \lambda) = f(x) + \varphi(y) + \lambda^\top g(x) + \lambda^\top h(x) + \frac{1}{2\mu} \|y - (Tx + \mu\lambda)\|^2 - \frac{\mu}{2} \|\lambda\|^2.$$

According to the proximal mapping defined in (2.2), the minimizer $y_{x, \lambda}^*$ of L^μ with respect to y is

$$y_{x, \lambda}^* = \text{prox}_{\mu\varphi}(Tx + \mu\lambda). \quad (4.3)$$

Evaluation of $L^\mu(x, y; \lambda, \lambda, \lambda)$ at this minimizer gives the following PAL

$$\mathcal{L}_\mu(x; \lambda, \lambda, \lambda) = f(x) + \varphi_\mu(Tx + \mu\lambda) + \lambda^\top g(x) + \lambda^\top h(x) - \frac{\mu}{2} \|\lambda\|^2. \quad (4.4)$$

By definition (2.1), this function is exactly the Moreau envelope of L^μ and thus is continuously differentiable with respect to both x and $(\lambda, \lambda, \lambda)$.

The PAL was proposed in [24], but without considering optimization constraints. When such constraints are taken into account, a recent study [68] identifies three fundamental challenges in extending PAL to settings with non-smooth objectives, with exponential stability analysis emerging as a critical theoretical obstacle. To prove exponential stability, the IQC framework [29, 45] provides systematic tools for exponential convergence analysis via linear matrix inequalities for systems in feedback interconnections involving linear dynamics and IQC-satisfying nonlinearities. While gradient and proximal mappings individually comply with IQC criteria, their compositions with nonlinear operators generally violate such constraints unless combined with affine operations [29]. Consequently, existing IQC-based PAL analyses for exponential convergence predominantly address equality constraints [24, 26, 36, 37, 68] or affine inequalities [5, 35, 57], leaving nonlinear convex inequality constraints unexplored. While some existing works address general constraints, they either guarantee only asymptotic convergence [33] or achieve semi-global exponential convergence [64]. To handle general optimization constraints while ensuring exponential convergence, we bypass the IQC framework. Instead, we establish in the following lemma that the PAL $\mathcal{L}_\mu(x; \lambda, \lambda, \lambda)$ is strongly convex in the primal variable x and strongly concave in the dual variable $(\lambda, \lambda, \lambda)$, a vital property for proving exponential convergence of our optimization algorithm.

Lemma 4.1. Under Assumptions 1 and 2, the PAL defined in (4.4) is α -strongly convex in x and $\left(\frac{\mu\ell}{\mu+\ell} + 2\mu\right)$ -strongly concave in $(\lambda, \lambda, \lambda)$.

Proof. Obviously L^μ in (4.1) is convex in (x, y) since it is a linear combination of the convex and affine functions $f(x)$, $\varphi(y)$, $g(x)$, $h(x)$, $Tx - y$ and $\|Tx - y\|^2$ (the convexity of the last one can be shown by a simple calculation) with the combination coefficients for convex functions being non-negative; furthermore, L^μ is strongly convex in x since $f(x)$ is strongly convex due to Assumption 1.

We now show that $\mathcal{L}_\mu(x; \lambda, \lambda, \lambda)$ is strongly concave in $(\lambda, \lambda, \lambda)$. Note

$$\begin{aligned} \mathcal{L}_\mu(x; \lambda, \lambda, \lambda) &= f(x) + \lambda^\top g(x) + \lambda^\top h(x) - \frac{\mu}{2} \|\lambda\|^2 \\ &\quad - \sup_{y \in \mathbb{R}^m} \left\{ -\varphi(y) - \frac{1}{2\mu} \|y - (Tx + \mu\lambda)\|^2 \right\} \\ &= f(x) + \lambda^\top g(x) + \lambda^\top h(x) - \frac{\mu}{2} \|\lambda\|^2 - \frac{1}{2\mu} \|Tx + \mu\lambda\|^2 \\ &\quad - \sup_{y \in \mathbb{R}^m} \left\{ -\varphi(y) - \frac{1}{2\mu} \|y\|^2 + y^\top \left(\frac{1}{\mu} Tx + \lambda \right) \right\} \\ &= f(x) + \lambda^\top g(x) + \lambda^\top h(x) - \frac{\mu}{2} \|\lambda\|^2 - \frac{\mu}{2} \left\| \frac{Tx}{\mu} + \lambda \right\|^2 - \tilde{\varphi}^* \left(\frac{1}{\mu} Tx + \lambda \right), \end{aligned}$$

where $\tilde{\varphi}^*$ is the Fenchel conjugate of the function $\tilde{\varphi}(y) = \varphi(y) + \frac{1}{2\mu}\|y\|^2$. Noting that $\tilde{\varphi}$ is $1/\mu$ -strongly convex, it then follows from [40, Theorem 4.2.2] that $\tilde{\varphi}^*$ is μ -smooth. Also, in view of the $1/\ell$ -Lipschitz continuous of φ assumed in Assumption 2 or equivalently the $(1/\ell + 1/\mu)$ -Lipschitz continuous of $\tilde{\varphi}$, one has $\frac{\mu\ell}{\mu+\ell}$ -strong convexity of φ^* . Therefore, $\tilde{\varphi}^*(\frac{1}{\mu}Tx + \tilde{\lambda})$ is a $\frac{\mu\ell}{\mu+\ell}$ -strongly convex function of $\tilde{\lambda}$. Also, both $\frac{\mu}{2}\|\tilde{\lambda}\|^2$ and $\frac{\mu}{2}\|\frac{Tx}{\mu} + \tilde{\lambda}\|^2$ are μ -strongly convex functions of $\tilde{\lambda}$. Therefore, $\mathcal{L}_\mu(x; \lambda, \tilde{\lambda}, \tilde{\lambda})$ is a $(\frac{\mu\ell}{\mu+\ell} + 2\mu)$ -strongly concave function of $(\lambda, \tilde{\lambda}, \tilde{\lambda})$. $\square$

With the above lemma, we are directed toward utilizing $\mathcal{L}_\mu(x; \lambda, \tilde{\lambda}, \tilde{\lambda})$ as a Lagrangian to design corresponding primal-dual dynamics. A preliminary work should be done to relate the optimal solution x^* of (3.1) to the saddle point of $\mathcal{L}_\mu(x; \lambda, \tilde{\lambda}, \tilde{\lambda})$. This is shown in the following lemma.

Lemma 4.2. Consider the optimization problem (3.1) under Assumptions 1–3.

- (1) Suppose there exists a point $x^* \in \mathbb{R}^n$ and multipliers $(\lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$ such that

$$\mathcal{L}_\mu(x^*; \lambda, \tilde{\lambda}, \tilde{\lambda}) \leq \mathcal{L}_\mu(x^*; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \leq \mathcal{L}_\mu(x; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \quad (4.5)$$

holds for all $(x; \lambda, \tilde{\lambda}, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$. Then x^* is an optimal solution to the optimization problem (3.1).

- (2) Suppose $x^* \in \mathbb{R}^n$ is an optimal solution to the optimization problem (3.1). Then there exist $(\lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$ such that (4.5) holds.

Proof. (1) We show that $(x^*, y^*; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*)$ with $y^* = y_{x^*, \tilde{\lambda}^*}^*$ defined in (4.3) satisfies the $\mathcal{KKT}$ system (3.3) so that (x^*, y^*) is an optimal solution to (3.2) and consequently x^* is an optimal solution to (3.1).

Firstly, we use the first inequality in (4.5), i. e.,

$$\begin{aligned} f(x^*) + \varphi(y_{x^*, \tilde{\lambda}^*}^*) + \lambda^{*\top} g(x^*) + \tilde{\lambda}^{*\top} h(x^*) + \tilde{\lambda}^{*\top} (Tx^* - y_{x^*, \tilde{\lambda}^*}^*) + \frac{1}{2\mu} \|Tx^* - y_{x^*, \tilde{\lambda}^*}^*\|^2 \\ \leq \mathcal{L}_\mu(x^*; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*). \end{aligned}$$

Noting that the right side of the above inequality is a constant and this inequality holds for all $\lambda \succeq 0$, we must have $g(x^*) \preceq 0$. Similarly, this inequality holds for all $\tilde{\lambda}$ so that one must have $h(x^*) = 0$. Also, this inequality holds for all $\tilde{\lambda}$, implying $Tx^* = y_{x^*, \tilde{\lambda}^*}^*$.

On the one hand, taking $(\lambda, \tilde{\lambda}) = (0, \tilde{\lambda}^*)$ in the first inequality in (4.5) gives $\lambda^{*\top} g(x^*) \geq 0$; on the other hand, $\lambda^* \succeq 0$ and $g(x^*) \preceq 0$ give $\lambda^{*\top} g(x^*) \leq 0$. Therefore, $g(x^*) \preceq 0$ gives $\lambda^{*\top} g(x^*) = 0$.

Lastly, we prove $(x^*; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*)$ satisfies (3.3f). With $y^* = y_{x^*, \tilde{\lambda}^*}^*$ and by referring to the right hand side of the inequality (4.5), one has

$$\begin{aligned} L^\mu(x^*, y_{x^*, \tilde{\lambda}^*}^*; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \\ = \mathcal{L}_\mu(x^*; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \leq \mathcal{L}_\mu(x; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*) \leq L^\mu(x, y_{x, \tilde{\lambda}^*}^*; \lambda^*, \tilde{\lambda}^*, \tilde{\lambda}^*). \end{aligned}$$

This means that x^* is a minimizer of $x \rightarrow L^\mu(x, y_{x^*, \lambda^*}^*; \lambda^*, \lambda^*, \lambda^*)$ so that its subdifferential set includes 0. Noting that $L^\mu(x^*, y; \lambda^*, \lambda^*, \lambda^*)$ achieves its minimum at y_{x^*, λ^*}^* , its derivative with respect to y at y_{x^*, λ^*}^* is zero. Now the necessary conditions for x^* to be a minimizer of $x \rightarrow L^\mu(x, y_{x^*, \lambda^*}^*; \lambda^*, \lambda^*, \lambda^*)$ becomes

$$0 = \nabla f(x^*) + \sum_{i=1}^r \lambda^{*\top} \nabla g_i(x^*) + \sum_{i=1}^s \lambda^{*\top} \nabla h_i(x^*) + T^\top \lambda^* + \frac{1}{\mu} T^\top (Tx^* - y_{x^*, \lambda^*}^*) \quad (4.6)$$

Also noting that y_{x^*, λ^*}^* is a minimizer of the function $y \rightarrow L^\mu(x^*, y; \lambda^*, \lambda^*, \lambda^*)$, the following optimality condition holds

$$\partial \varphi(y_{x^*, \lambda^*}^*) - \lambda^* - \frac{1}{\mu} (Tx^* - y_{x^*, \lambda^*}^*) \ni 0 \quad (4.7)$$

These two results (4.6) and (4.7), together with $y_{x^*, \lambda^*}^* = Tx^*$, give

$$\nabla f(x^*) + \sum_{i=1}^r \lambda^{*\top} \nabla g_i(x^*) + \sum_{i=1}^s \lambda^{*\top} \nabla h_i(x^*) + T^\top \lambda^* p \ni 0.$$

Therefore, $(x^*; \lambda^*, \lambda^*, \lambda^*)$ satisfies the KKT system (3.3), implying that (x^*, y^*) is an optimal solution of (3.2) and consequently x^* is an optimal solution of (3.1).

(2) Suppose $x^* \in \mathbb{R}^n$ is an optimal solution to the optimization problem (3.1) so that (x^*, y^*) with $y^* = Tx^*$ is an optimal solution to (3.2), implying the existence of $(\lambda^*, \lambda^*, \lambda^*)$ such that $(x^*, y^*; \lambda^*, \lambda^*, \lambda^*)$ satisfies the KKT system (3.3). Therefore,

$$\mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*) \leq L^\mu(x^*, y^*; \lambda^*, \lambda^*, \lambda^*) = f(x^*) + \varphi(y^*). \quad (4.8)$$

On the other hand, the strong duality for the problem (3.2) implies

$$\begin{aligned} f(x^*) + \varphi(y^*) &= \min_{x, y} \left\{ f(x) + \varphi(y) + \lambda^{*\top} g(x) + \lambda^{*\top} h(x) + \lambda^{*\top} (Tx - y) \right\} \\ &\leq f(x) + \varphi(y) + \lambda^{*\top} g(x) + \lambda^{*\top} h(x) + \lambda^{*\top} (Tx - y) \\ &\leq \mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*). \end{aligned} \quad (4.9)$$

Combining the inequalities (4.8) and (4.9), one obtains $\mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*) \leq \mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*)$ which is the second inequality in (4.5).

To prove the first inequality in (4.5), we recall the KKT systems (3.3) and have the following inequality

$$\begin{aligned} \mathcal{L}_\mu(x^*; \lambda, \lambda, \lambda) &= f(x^*) + \varphi(y_{x^*, \lambda}^*) + \lambda^\top g(x^*) + \lambda^\top h(x^*) + \lambda^\top (Tx^* - y_{x^*, \lambda}^*) \\ &\quad + (1/2\mu) \|Tx^* - y_{x^*, \lambda}^*\|^2 \\ &= f(x^*) + \varphi(y_{x^*, \lambda}^*) + \lambda^{*\top} g(x^*) + \lambda^{*\top} h(x^*) + \lambda^{*\top} (Tx^* - y_{x^*, \lambda}^*) \\ &\quad + (1/2\mu) \|Tx^* - y_{x^*, \lambda}^*\|^2 \\ &= \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*). \end{aligned}$$

This completes the second part of the proof. $\square$

4.2. Projection-free primal-dual optimization dynamics

We are now in a position to design a primal-dual dynamics building on the PAL (4.4) to seek the optimal solution (x^*, y^*) of the optimization problem (3.2) with exponential convergence. This algorithm excuses gradient descent on x and gradient ascent on $(\lambda, \bar{\lambda}, \bar{\lambda})$. However, the gradient descent of $\mathcal{L}_\mu$ on λ requires a projection onto the positive quadrant to keep the resulting λ dynamics stay non-negative; more specifically, $\dot{\lambda} = [\nabla_\lambda \mathcal{L}_\mu]_+$, which has a discontinuous right-hand side. In this paper, we propose a new design procedure for this Lagrangian dynamics without using projection but referring to mirror ascent; it comes with the advantage that the resulting dynamics is smooth so that its convergence analysis and numerical simulations are relatively simple compared with non-smooth projection based dynamics. More specifically, we propose the following projection-free algorithm which, for the PAL (4.4), performs gradient descent on the primal variable x , gradient ascent on the dual variables $(\lambda, \bar{\lambda})$, and mirror ascent on the dual variable λ , specified as follows,

$$\begin{cases} \dot{x} = -\nabla f(x) - T^\top \nabla \varphi_\mu(Tx + \mu \bar{\lambda}) - \lambda^\top \nabla g(x) - \bar{\lambda}^\top \nabla h(x), & (4.10a) \\ \dot{\lambda} = [\lambda \odot (1 + \eta \odot \lambda)] \odot g(x), \lambda(0) \geq 0, & (4.10b) \\ \dot{\lambda} = h(x), & (4.10c) \\ \dot{\bar{\lambda}} = \mu \nabla \varphi_\mu(Tx + \mu \bar{\lambda}) - \mu \bar{\lambda}. & (4.10d) \end{cases}$$

Remark 4.3. Some remarks are given regarding the intuition behind the projection-free dynamics (4.10b). Recall that the mirror descent algorithm is devoted to the constrained minimization problem $\min_{y \in \mathcal{Y}} F(y)$, where $\mathcal{Y}$ is a convex set in an Euclidian space. Let $\phi : \mathcal{Y} \rightarrow \mathbb{R}$ be a function which is strictly convex and twice differentiable. Its conjugate convex function $\phi^* : \mathcal{Y} \rightarrow \mathbb{R}$ is defined as $\phi^*(\omega) = \sup_{y \in \mathcal{Y}} \{(\omega, y) - \phi(y)\}$, which can be shown as convex and differentiable [59, Theorem 26.3]. Let $\mathcal{Z} = \{z | z = \nabla \phi(y), y \in \mathcal{Y}\}$ be the image of $\mathcal{Y}$ under the mapping $\nabla \phi$. Then $\nabla \phi : \mathcal{Y} \rightarrow \mathcal{Z}$ and $\nabla \phi^* : \mathcal{Z} \rightarrow \mathcal{Y}$. The continuous-time mirror descent algorithm for the above constrained minimization problem takes the following form ([58, e.q. (5)]): $\dot{z} = -\nabla F(y), y = \nabla \phi^*(z)$. The second equation $y = \nabla \phi^*(z)$ is equivalent to $z = \nabla \phi(y)$. Taking derivative yields $\dot{z} = \nabla^2 \phi(y) \dot{y}$. Therefore the first equation in the mirror descent becomes $\dot{y} = -[\nabla^2 \phi(y)]^{-1} \nabla F(y)$. Choosing the constraint set as $\mathcal{Y} = \{y | y_i \geq 0\}$ and $\phi(y) = \frac{\eta}{2} \|y\|^2 + \sum_{i=1}^n y_i \ln y_i$ which is well defined on $\mathcal{Y}$, the above mirror dynamics can be written as $\dot{y} = -\text{diag}[y_i / (1 + \eta y_i)]_{i=1}^n \nabla F(y)$. Letting $y = \lambda_i$ and $F(\lambda_i)$ be a function of λ_i defined as the right hand side of equation (4.4), then the above dynamics is exactly (4.10b).

Before carrying out the convergence analysis of the algorithm (4.10), the relationship between the equilibrium of (4.10) and the optimal solution of (3.1) is revealed in the following lemma.

Lemma 4.4. A point $(x^*; \lambda^*, \bar{\lambda}^*, \bar{\lambda}^*) \in \mathbb{R}^n \times \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$ is the equilibrium of (4.10) if and only if x^* is the optimal solution to the optimization problem (3.2).

Proof. Suppose x^* is an optimal solution. It follows from Lemma 4.2 that there exists $(\lambda^*, \lambda^*, \lambda^*) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$ such that the saddle point inequality (4.5) holds. This implies that at $(\lambda^*, \lambda^*, \lambda^*)$ it holds $\partial \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)/\partial x = 0$, $\partial \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)/\partial \lambda = 0$ and $\partial \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)/\partial \lambda = 0$. These three equations exactly shows that the right hands of the equations (4.10a), (4.10c), and (4.10d) vanish at $(\lambda^*, \lambda^*, \lambda^*)$. The right hand of the equation (4.10b) also takes zero value at $(\lambda^*, \lambda^*, \lambda^*)$ in view of KKT subsystem (3.3e). Therefore, $(x^*; \lambda^*, \lambda^*, \lambda^*)$ is the equilibrium of (4.10).

On the other hand, letting $(x^*; \lambda^*, \lambda^*, \lambda^*)$ be the equilibrium of (4.10) and inserting it into (4.10a), (4.10c), and (4.10d), one obtains

$$\frac{\partial \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)}{\partial x} = 0, \quad (4.11)$$

$$\frac{\partial \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)}{\partial \lambda} = 0, \quad (4.12)$$

$$\frac{\partial \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)}{\partial \lambda} = 0. \quad (4.13)$$

Recall the fact that, for a differentiable convex (concave) function $F(x)$, a necessary and sufficient condition for $\bar{x}$ to be the minima (maxima) of F is $\nabla F(\bar{x}) = 0$. By using this fact, the result (4.11) gives the minima inequality

$$\mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*) \leq \mathcal{L}_\mu(x, \lambda^*, \lambda^*, \lambda^*), \quad (4.14)$$

and the results in (4.12)–(4.13) indicates the maxima inequality $\mathcal{L}_\mu(x^*; \lambda^*, \lambda, \lambda) \leq \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)$ with respect to (λ, λ) . Due to the decoupling of λ with (λ, λ) in the PAL (4.4), this inequality holds for all λ rather than only on λ^* ,

$$\mathcal{L}_\mu(x^*; \lambda, \lambda, \lambda) \leq \mathcal{L}_\mu(x^*; \lambda, \lambda^*, \lambda^*). \quad (4.15)$$

Inserting $(x^*; \lambda^*, \lambda^*, \lambda^*)$ into (4.10b) give $\lambda^* \odot g(x^*) = 0$; i.e., $\lambda_i^* g_i(x^*) = 0, i = 1, \dots, r$. This shows that $g_i(x^*) = 0$ or $\lambda_i^* = 0$. The former is nothing but $\partial \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)/\partial \lambda = 0$ so that the function $\lambda \rightarrow \mathcal{L}_\mu(x^*; \lambda, \lambda^*, \lambda^*)$ is maximized at λ^* ; the latter, together with $g_i(x^*) \leq 0$, shows that the function $\lambda \rightarrow \mathcal{L}_\mu(x^*; \lambda, \lambda^*, \lambda^*)$ is maximized at λ^* . Both cases indicate that

$$\mathcal{L}_\mu(x^*; \lambda, \lambda^*, \lambda^*) \leq \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*). \quad (4.16)$$

The combination of (4.15) and (4.16) gives

$$\mathcal{L}_\mu(x^*; \lambda, \lambda, \lambda) \leq \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*). \quad (4.17)$$

The two inequalities (4.14) and (4.17) tell us that $(x^*; \lambda^*, \lambda^*, \lambda^*)$ is the saddle point of $\mathcal{L}_\mu$. According to the first result in Lemma 4.2, x^* is an optimal solution to the optimization problem (3.2). $\square$

With the relationship between the equilibrium of the optimization dynamics (4.10) and the optimal solution to the problem (3.2) discovered in the above lemma, we have the following result which proves the exponential stability of our algorithm (4.10).

Theorem 4.5. For the optimization problem (3.2) under Assumptions 1–3, consider the optimization algorithm (4.10). Then, for any trajectory of (4.10) with initial conditions $x(0) \in \mathbb{R}^n$ and $(\lambda(0), \bar{\lambda}(0), \bar{\lambda}(0)) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$, the sub-trajectory $x(t)$ converges to the optimal solution x^* exponentially as $t \rightarrow +\infty$.

Proof. Construct a Lyapunov candidate $V(x; \lambda, \bar{\lambda}, \bar{\lambda}) = V_1 + V_2 + V_3 + V_4$ with

$$\begin{aligned} V_1 &= \frac{1}{2} \|x - x^*\|^2, \\ V_2 &= \frac{1}{2} \sum_{i=1}^r \eta_i (\lambda_i - \lambda_i^*)^2, \\ V_3 &= \sum_{i \in \Omega} D_\psi(\lambda_i, \lambda_i^*) + \sum_{i \notin \Omega} (\lambda_i - \lambda_i^*)^2, \\ V_4 &= \frac{1}{2} \sum_{i=1}^s (\bar{\lambda}_i - \bar{\lambda}_i^*)^2 + \frac{1}{2} \sum_{i=1}^s (\bar{\lambda}_i - \bar{\lambda}_i^*)^2, \end{aligned}$$

where $\Omega = \{i | \lambda_i^* \neq 0\}$ and $D_\psi(\lambda_i, \lambda_i^*) \geq 0$ is the Bregman divergence [15] between λ_i and λ_i^* with respect to $\psi(t) = t \ln t$. Therefore, $V \geq 0$, and also $V(x; \lambda, \bar{\lambda}, \bar{\lambda}) = 0$ if and only if $(x; \lambda, \bar{\lambda}, \bar{\lambda}) = (x^*; \lambda^*, \bar{\lambda}^*, \bar{\lambda}^*)$. We now use this candidate Lyapunov function to prove exponential stability of the equilibrium $(x^*; \lambda^*, \bar{\lambda}^*, \bar{\lambda}^*)$ of the system (4.10) by referring to the criterion in [42, Corollary 3.4]. To this end, we will show that V is lower and upper bounded by quadratic functions and the time derivative of V is upper bounded by a negative quadratic function. The following two parts are the details.

(1) We first show that V can be upper and lower bounded by quadratic functions of $(x; \lambda, \bar{\lambda}, \bar{\lambda})$ (we only prove this for V_3 since this is obvious for V_1, V_2, V_4). For $i \in \Omega$ which implies $\lambda_i^* > 0$, the Bregman $D_\psi(\lambda_i, \lambda_i^*)$ can be calculated as

$$D_\psi(\lambda_i, \lambda_i^*) = \psi(\lambda_i) - \psi(\lambda_i^*) - \psi'(\lambda_i^*)(\lambda_i - \lambda_i^*) = \frac{1}{2\varsigma_i} (\lambda_i - \lambda_i^*)^2.$$

with ς_i lying between λ_i^* and λ_i . In view of $\lambda_i^* > 0$ and $\lambda_i \rightarrow \lambda_i^*$, ς_i is a positive constant. Denoting $\tilde{a} = \min_{i \in \Omega} \{1/2\varsigma_i\}$ and $\tilde{b} = \max_{i \in \Omega} \{1/2\varsigma_i\}$, one has $\tilde{a} \leq D_\psi(\lambda_i, \lambda_i^*) \leq \tilde{b}$. With this, one can obtain the following lower and upper quadratic bounds for V_3 ,

$$\begin{aligned} a \| \lambda - \lambda^* \|^2 &\leq \sum_{j \in \Omega} \tilde{a} |\lambda_j - \lambda_j^*|^2 + \sum_{j \notin \Omega} (\lambda_j - \lambda_j^*)^2 \\ &\leq V_3 \leq \sum_{j \in \Omega} \tilde{b} |\lambda_j - \lambda_j^*|^2 + \sum_{j \notin \Omega} (\lambda_j - \lambda_j^*)^2 \leq b \| \lambda - \lambda^* \|^2, \end{aligned}$$

where $a = \min(\tilde{a}, 1) > 0$ and $b = \max(\tilde{b}, 1) > 0$. Therefore, the function V has quadratic lower and upper bounds.

(2) The time derivative of V along the trajectories of (4.10) is,

$$\dot{V}|_{(4.10)} = -(x - x^*)^\top \frac{\partial \mathcal{L}_\mu}{\partial x} + \sum_{i=1}^r \frac{\eta_i \lambda_i (\lambda_i - \lambda_i^*)}{1 + \eta_i \lambda_i} \frac{\partial \mathcal{L}_\mu}{\partial \lambda} + \sum_{i=1}^r \frac{\lambda_i}{1 + \eta_i \lambda_i} \frac{\partial \mathcal{L}_\mu}{\partial \lambda}$$

$$\begin{aligned}
& - \sum_{i \in \Omega} \frac{\lambda_i^*}{1 + \eta_i \lambda_i} \frac{\partial \mathcal{L}_\mu}{\partial \lambda} + \sum_{i=1}^s (\lambda_i - \lambda_i^*) \frac{\partial \bar{L}}{\partial \lambda} + \sum_{i=1}^m (\lambda_i - \lambda_i^*) \frac{\partial L_\mu}{\partial \lambda} \\
& = -(x - x^*)^\top \frac{\partial \mathcal{L}_\mu}{\partial x} + (\lambda - \lambda^*)^\top \frac{\partial \mathcal{L}_\mu}{\partial \lambda} + (\lambda - \lambda^*)^\top \frac{\partial \mathcal{L}_\mu}{\partial \lambda} + (\lambda - \lambda^*)^\top \frac{\partial \mathcal{L}_\mu}{\partial \lambda} \\
& = -(x - x^*)^\top \frac{\partial \mathcal{L}_\mu}{\partial x} + (\Lambda - \Lambda^*)^\top \frac{\partial \mathcal{L}_\mu}{\partial \Lambda},
\end{aligned}$$

where $\Lambda = (\lambda, \lambda, \lambda)$ and $\Lambda^* = (\lambda^*, \lambda^*, \lambda^*)$. Noting that $\mathcal{L}_\mu(x; \lambda, \lambda, \lambda)$ has been proved in Lemma 4.1 to be strongly convex in x and strongly concave in $(\lambda, \lambda, \lambda)$, the above inequality becomes

$$\begin{aligned}
\dot{V}|_{(4.10)} & \leq [\mathcal{L}_\mu(x^*; \lambda, \lambda, \lambda) - \mathcal{L}_\mu(x; \lambda, \lambda, \lambda)] + [\mathcal{L}_\mu(x; \lambda, \lambda, \lambda) - \mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*)] \\
& \quad - \alpha \|x - x^*\|^2 - \left(\frac{\mu \ell}{\mu + \ell} + 2\mu \right) \|\Lambda - \Lambda^*\|^2.
\end{aligned} \tag{4.18}$$

This shows that $\dot{V}$ has minus quadratic upper bound. $\square$

Usually, in designing algorithms for constrained optimization problem, either hard constraints or soft constraints can be considered. In the former, the optimization constraints should be satisfied on the fly (e.g., the interior algorithms), while in the latter, rather than forcing each transient state of the iterate to obey the constraints, one requires the optimization constraints to be satisfied only at the asymptotic state. Since the ultimate objective is to compute the optimal solution that satisfies the constraints, our algorithm just follows the second line of soft constraints and therefore there is no need in this paper to show positive feasibility for each transient state $\lambda(t)$ but feasibility for the asymptotic state of $\lambda(t)$ only. This is indeed the case in view of the equilibrium analysis in Lemma 4.4.

Remark 4.6. We note that our method is capable of achieving fully smooth dynamics while handling general convex inequality constraints. In primal-dual optimization dynamics based on the PAL framework, the multiplier λ associated with inequality constraint must remain nonnegative. Existing methods enforce this by incorporating a projection onto the positive quadrant [30, 33, 64], but this introduces nonsmooth dynamics that complicate stability analysis. To address this, smooth alternatives have been proposed to eliminate projection. The works [2, 11, 57] handle inequality constraints without explicit projection, instead they replace projection-based updates with max-based regularization (specifically algorithm (9a)–(9c) in [2], (9a)–(9b) in [57], and (1b) in [11]), implicitly ensuring nonnegativity of the multiplier associated with inequality constraint. While this improves computational simplicity, the max operator still makes the optimization dynamics nonsmooth in the classical sense. In contrast, our projection-free dynamics (4.10b) is fully smooth, and this smooth nature enables rigorous stability certification via classical Lyapunov techniques, bypassing the need for nonsmooth analysis tools like differential inclusions. More important, the max-based method in [2, 57] are limited to affine inequality constraints, with no extension to general convex inequality constraints. Our approach overcomes this limitation and extends to general convex inequality constraints.

Remark 4.7. Asides from its ability to obtain fully smooth optimization dynamics and deal with general convex inequality constraints, our method can yield exponential convergence rather than asymptotic convergence. As noted in [68], the combination of general convex constraints and nonsmooth objectives poses a significant challenge to achieving exponential stability. For instance, existing studies on general inequality constraints typically ensure only asymptotic convergence [33] or semi-global exponential convergence [64]. IQC is a widely used framework for exponential convergence analysis. However, existing IQC-based approaches primarily address affine-constrained settings, such as equality constraints [24, 26, 36, 37, 68] or affine linear inequalities [?], leaving nonlinear convex inequality constraints unexamined. Similar limitations appear in projection-free methods [2] and exponential stability analyses [57], which remain restricted to affine inequalities. We do not use the IQC method, we can prove exponential stability without assuming affine linearity assumption on inequality constraints.

Note that, in the proof of Theorem 4.5, the strong concavity of the PAL function $\mathcal{L}_\mu$ on $(\lambda, \lambda^*, \lambda^*)$ depends only on the Lipschitz continuous of φ in the sense of Assumption 2, but not on the strong convexity of f . If the strong convexity assumption on f made in Assumption 1 is weakened to be convex, the PAL $\mathcal{L}_\mu$ in (4.4) is only convex on x , but still strongly concave on $(\lambda, \lambda^*, \lambda^*)$. In this case, only asymptotic stability of the optimization dynamics (4.10) can be ensured. Before presenting the result in this case, the following assumption is given.

Assumption 1'. The function f is convex and continuously differentiable.

Theorem 4.8. For the optimization problem (3.2) under Assumptions 1', 2 and 3, consider the optimization algorithm (4.10). Then, for any trajectory of (4.10) with initial conditions $x(0) \in \mathbb{R}^n$ and $(\lambda(0), \lambda^*(0), \lambda^*(0)) \in \mathbb{R}_+^r \times \mathbb{R}^s \times \mathbb{R}^m$, the sub-trajectory $x(t)$ converges to the optimal solution x^* asymptotically as $t \rightarrow +\infty$.

Proof. Consider also the Lyapunov function constructed in the proof of Theorem 4.5. The time derivative of V calculated in (4.18) is now modified as

$$\begin{aligned} \dot{V}|_{(4.10)} &\leq [\mathcal{L}_\mu(x^*; \lambda, \lambda^*, \lambda^*) - \mathcal{L}_\mu(x; \lambda, \lambda^*, \lambda^*)] \\ &\quad + [\mathcal{L}_\mu(x; \lambda, \lambda^*, \lambda^*) - \mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*)] - \left(\frac{\mu\ell}{\mu + \ell} + 2\mu \right) \|\Lambda - \Lambda^*\|^2. \\ &\leq [\mathcal{L}_\mu(x^*; \lambda, \lambda^*, \lambda^*) - \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)] \\ &\quad + [\mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*) - \mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*)] - \left(\frac{\mu\ell}{\mu + \ell} + 2\mu \right) \|\Lambda - \Lambda^*\|^2. \end{aligned} \quad (4.19)$$

Note that the x -related quadratic form in the right hand side of (4.18) is dropped due to the lack of strong convexity of $\mathcal{L}_\mu$. Since $(x^*; \lambda^*, \lambda^*, \lambda^*)$ is a saddle point of $\mathcal{L}_\mu$, the two curl brackets are both nonnegative. Therefore, the right hand side of (4.19) is less than or equal to zero. Thus $\dot{V} \leq 0$.

We now apply the LaSalle's invariance principle to prove the asymptotic stability; that is, we only need to prove that $\dot{V} = 0$ implies $(x; \lambda, \lambda^*, \lambda^*) = (x^*; \lambda^*, \lambda^*, \lambda^*)$. To this

end, letting $\dot{V} = 0$ gives $\Lambda = \Lambda^*$ and $\mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*) = \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)$. Noting again $\Lambda = \Lambda^*$, one has

$$\mathcal{L}_\mu(x; \lambda, \lambda, \lambda) = \mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*) = \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*) \leq \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*),$$

or $\mathcal{L}_\mu(x; \lambda, \lambda, \lambda) \leq \mathcal{L}_\mu(x; \lambda^*, \lambda^*, \lambda^*) \leq \mathcal{L}_\mu(x^*; \lambda^*, \lambda^*, \lambda^*)$. This shows that $(x; \lambda^*, \lambda^*, \lambda^*)$ is a saddle point of $\mathcal{L}_\mu$ and therefore x is the optimal solution of the problem (3.2) in view of the first result in Lemma 4.2. By uniqueness of the optimal solution, we have $x = x^*$. Thus, $(x; \lambda, \lambda, \lambda) = (x^*; \lambda^*, \lambda^*, \lambda^*)$. As a consequence of LaSalle's invariance principle, the equilibrium $(x^*; \lambda^*, \lambda^*, \lambda^*)$ is asymptotically stable. $\square$

Remark 4.9. Although small values of the regularization parameter μ would cause slow convergence, this slowness can be alleviated via the “continuation schemes” carried out in the following way. Firstly, given a particular parameter μ , one runs the algorithm to obtain a solution. With this solution as the “warm-start” and taking it as the initial state of the second round of execution of the optimization dynamics with a smaller parameter μ , one obtains a second solution. Continuing this procedure with a sequence of decreasing μ , the optimal solution can be computed in succession. This kind of “continuation schemes” have been found quite effective in speeding up the algorithm [3, 31].

5. SIMULATION EXAMPLE

As an example, we applied our algorithm to the following distributed optimization problem on a network with N nodes,

$$\begin{cases} \text{minimize} & \sum_{i=1}^N f_i(x), \\ \text{subject to} & g_i(x) \preceq 0, \\ & h_i(x) = 0, i = 1, \dots, N. \end{cases} \quad (5.1)$$

Here $x \in \mathbb{R}^n$ is the decision variable, $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is the local cost function on node i , $g_i = (g_{i1}, \dots, g_{ir_i})^\top : \mathbb{R}^n \rightarrow \mathbb{R}^{r_i}$ is the inequality constrained function on node i , and $h_i = (h_{i1}, \dots, h_{is_i})^\top : \mathbb{R}^n \rightarrow \mathbb{R}^{s_i}$ is the equality constrained function on node i . If there is no constraints for agent i , one simply sets corresponding constraint functions to be zero.

In view of large N , it is difficult for a single agent to do the optimization task. Instead, one usually uses N agents to cooperatively solve the problem, where each agent i executes a subtask of local optimization by only using information from agent i (i.e., the information f_i, g_i, h_i as well as their gradients) and those information from its neighbors $\mathcal{N}_i$. The neighboring relationship encodes the cooperation among these agents and is described by a graph. The cooperation makes local optimization algorithms compute the optimal solution in a consensus way. That is, letting $x_i \in \mathbb{R}^n$ denote the estimation of the optimal solution x^* by local algorithm of agent $i \in \{1, \dots, N\}$, the cooperation among these agents should be directed toward rendering $x_1 = x_2 = \dots = x_N = x^*$. Denoting respectively the graph Laplacian and incidence matrix of the graph by L and T so that $L = T^\top T$, the consensus $x_1 = x_2 = \dots = x_N$ among agents can be enforce by imposing the equality constraint $(T \otimes I_n)\mathbf{x} = 0$ with $\mathbf{x} = \text{col}(x_1, \dots, x_n) \in \mathbb{R}^{nN}$ if the

graph is connected. By adding to the objective function in (5.1) the following penalty function

$$\varphi(T\mathbf{x}) = \begin{cases} 0, & T\mathbf{x} = 0 \\ +\infty, & \text{otherwise,} \end{cases} \quad (5.2)$$

the optimization problem (5.1) can be represented as the following form

$$\begin{cases} \text{minimize} & \sum_{i=1}^N f_i(x_i) + \varphi(T\mathbf{x}), \\ \text{subject to} & g_i(x_i) \leq 0, \\ & h_i(x_i) = 0, i = 1, \dots, N. \end{cases} \quad (5.3)$$

Defining the objective function $F(\mathbf{x}) = \sum_{i=1}^N f_i(x_i)$ and stack constraint functions $G(\mathbf{x}) = \text{col}[g_1(x_1), \dots, g_N(x_N)]$, $H(\mathbf{x}) = \text{col}[h_1(x_1), \dots, h_N(x_N)]$, the problem (5.3) can be written in a compact form as

$$\begin{cases} \text{minimize} & F(\mathbf{x}) + \varphi(T\mathbf{x}), \\ \text{subject to} & G(\mathbf{x}) \leq 0, \\ & H(\mathbf{x}) = 0. \end{cases} \quad (5.4)$$

We now apply the algorithm (4.10) to the nonsmooth composite optimization problem (5.4). Note that, for the function φ defined in (5.2), it is easy to show that $\nabla\varphi(y) = \frac{1}{\mu}y$. Furthermore, corresponding to the multipliers λ and λ in (4.10) for a single agent, we use $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ir_i})^\top$ and $\lambda_i = (\lambda_{i1}, \dots, \lambda_{is_i})^\top$ to respectively denote multipliers corresponding to constraints $g_i(x) \leq 0$ and $h_i(x) = 0$; more specifically λ_{ij} is the multiplier for $g_{ij}(x) \leq 0$ for $j = 1, \dots, r_i$ and λ_{ij} is the multiplier for $h_{ij}(x) = 0$ for $j = 1, \dots, s_i$. Note that if there is no inequality or equality constraint for agent i , we simply set λ_i or λ_i to be zero vector of appropriate dimension. As for the multiplier λ (bold lambda used here to account for multiple agents) for the constraints $T\mathbf{x} = \mathbf{y}$ in (4.10), we introduce a new multiplier $\lambda' = T\lambda$ with partition $\lambda' = \text{col}[\lambda'_1, \dots, \lambda'_N]$. With this new multiplier, the subequations (4.10a) and (4.10d) have a simpler forms. More specifically, the resulting optimization algorithm has the following distributed form

$$\begin{cases} \dot{x}_i = \frac{1}{\mu} \sum_{j \in \mathcal{N}_i} (x_j - x_i) - \nabla f_i(x_i) - \sum_{j=1}^{r_i} \lambda_{ij} \nabla g_{ij}(x_i) - \sum_{j=1}^{s_i} \lambda_{ij} \nabla h_{ij}(x_i) - \lambda'_i \end{cases} \quad (5.5a)$$

$$\begin{cases} \dot{\lambda}_{ij} = \frac{\lambda_{ij}}{1 + \eta_{ij} \lambda_{ij}} g_{ij}(x_i), \quad j = 1, \dots, r_i, \end{cases} \quad (5.5b)$$

$$\begin{cases} \dot{\lambda}_{ij} = h_{ij}(x_i), \quad j = 1, \dots, s_i, \end{cases} \quad (5.5c)$$

$$\begin{cases} \dot{\lambda}'_i = - \sum_{j \in \mathcal{N}_i} (x_j - x_i), \end{cases} \quad (5.5d)$$

where $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$ is the neighboring agents of agent i . This algorithm is distributed since the optimization dynamics for agent i depends on information from its neighboring agents. See also [24] for distributed implementation without considering optimization constraints.

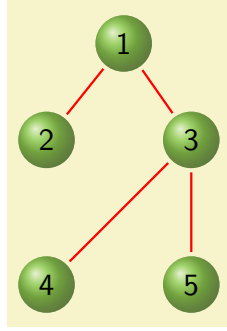


Fig. 1. The topology of five agents described by a graph.

We now apply the above distributed optimization algorithm to the following Rosen–Suzuki problem [18]

$$\begin{aligned}
 \min \quad & x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\
 \text{s.t.} \quad & -8 + x_1 - x_2 + x_3 - x_4 + x_1^2 + x_2^2 + x_3^2 + x_4^2 \leq 0, \\
 & -10 - x_1 - x_4 + x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 \leq 0, \\
 & -5 + 2x_1 - x_2 - x_4 + 2x_1^2 + x_2^2 + x_3^2 = 0.
 \end{aligned}$$

The optimal solution is shown in [18] to be $x^* = (0, 1, 2, -1)$.

We now use our distributed algorithm (5.5) to recover this result. Note the function corresponding to the equality constraint is not affine which is not satisfied by our theorems. However, since our theorems only give sufficient conditions, this means that our algorithms may work even if these sufficient conditions are not satisfied. Our simulation shows that this is indeed the case for the above example. To run the simulation, we use five agents whose coupling is described by the graph in Figure 1. The state of each agent lies in $\mathbb{R}^4$. To run simulation under this setup, define $x = (x_1, x_2, x_3, x_4)^\top$ and let $f_1(x) = x_1^2 + x_2^2$, $f_2(x) = 2x_3^2 + x_4^2$, $f_3(x) = -5x_1 - 5x_2$, $f_4(x) = -21x_3$, $f_5(x) = 7x_4$, $g_1(x) = -8 + x_1 - x_2 + x_3 - x_4 + x_1^2 + x_2^2 + x_3^2 + x_4^2$, $g_2(x) = -10 - x_1 - x_4 + x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2$, $h_1(x) = -5 + 2x_1 - x_2 - x_4 + 2x_1^2 + x_2^2 + x_3^2$. We assume that agent 1 has only access to f_1, g_1, h_1 , agent 2 has only access to f_2, g_2 , and agent i has only access to f_i for $i = 3, 4, 5$. The parameters η_{ij} in (5.5) is only required to be positive so that we chose them to be 1 for briefly. The initial states of the five agents are $x_1(0) = (3, 4, -3, 4)^\top$, $x_2(0) = (1, -2, 4, 2)^\top$, $x_3(0) = (-3, -4, 3, 3)^\top$, $x_4(0) = (3, 1, 2, -3)^\top$, $x_5(0) = (4, -2, -4, 1)^\top$, the initial states for the multipliers corresponding to the inequality constraints of agents 1 and 2 are given as $\lambda_1(0) = \lambda_2(0) = 3$, the initial state for the multiplier corresponding to the equality constraints of agents 1 is given as $\lambda_1(0) = 3$, the initial states for the transformed multiplier λ_i of the five agents are given as $\lambda_1(0) = \dots = \lambda_5(0) = (1, 2, 3, 4)^\top$. The simulation result, shown in Figure 2, indicates that the first, second, third, and fourth components of five agents converge to 0, 1, 2, -1 respectively. To save place, the time evolutions of the multipliers are not plotted here.

Obviously, we can also use, say 7 agents, to simulate the optimal solution. Which

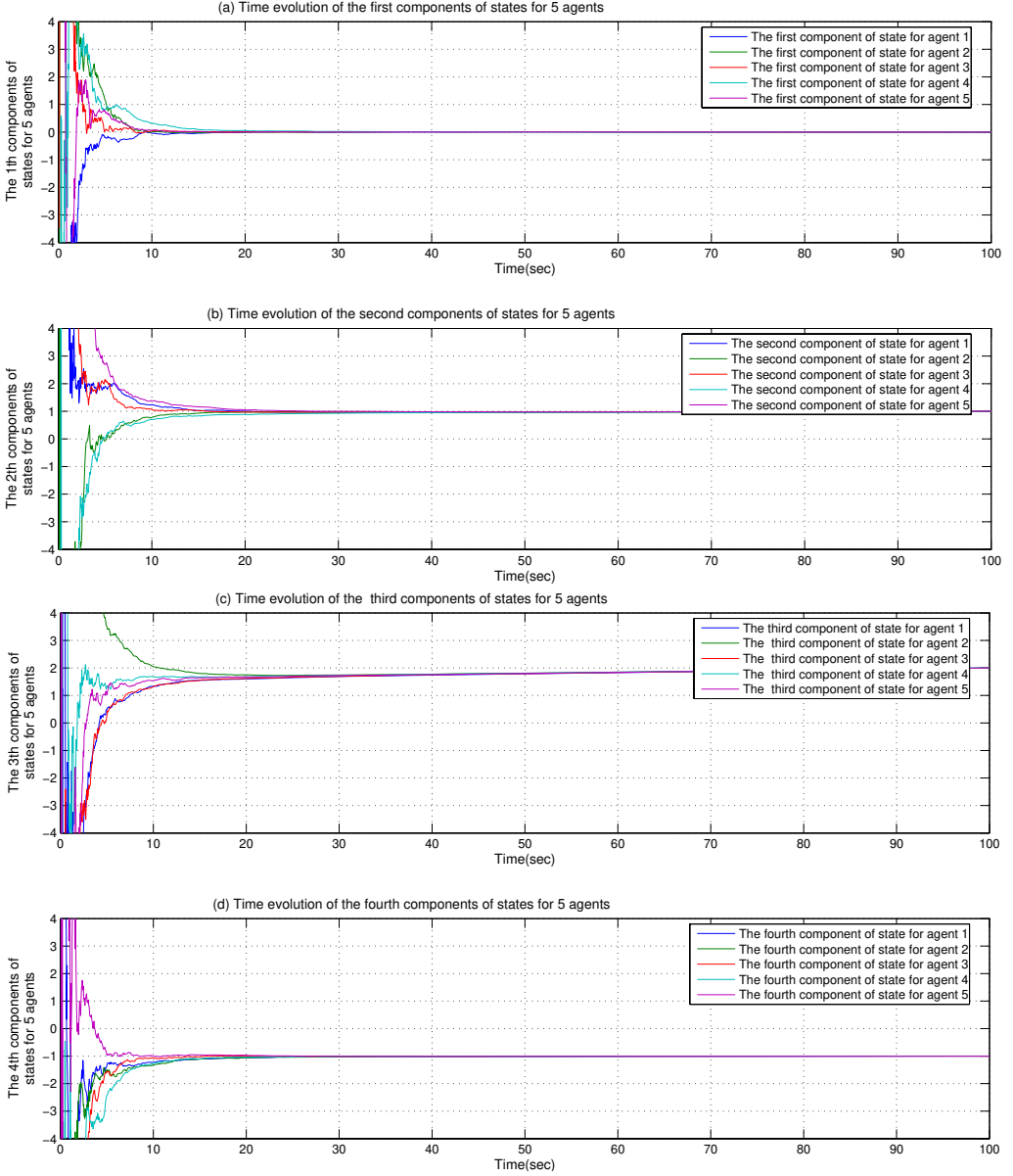


Fig. 2. The time evolution of the states for 5 agents. (a) The first components of the five agents all converge to 0; (b) The second components of the five agents all converge to 1; (c) The third components of the five agents all converge to 2; (d) The fourth components of the five agents all converge to -1. Therefore, each state of the 5 agents converges to the optimal solution $(0, 1, 2, -1)$.

number to use is application depend. Due to space limitation, the simulation result is omitted here.

6. CONCLUSIONS

This paper introduces a projection-free optimization dynamics for nonsmooth composite optimization problems involving both equality and convex inequality constraints. Unlike conventional projection-based approaches, which introduce nonsmooth dynamics and involve complicate stability analysis, the proposed method avoids using projection and ensures a fully smooth optimization process. This smoothness allows for rigorous stability analysis using classical Lyapunov techniques, eliminating the need for nonsmooth analysis tools such as differential inclusions.

Although several studies have also attempted to design constrained optimization algorithms without explicitly relying on projection methods, they instead substitute the projection operator with the relatively easier-to-compute max operator, still rendering the resulting optimization dynamics nonsmooth in the classical sense. Our optimization dynamics, which operates without projection, is entirely smooth and facilitates stability analysis through traditional Lyapunov methods, eliminating the necessity for nonsmooth techniques such as differential inclusions. More importantly, unlike the max-based approach which is confined to affine inequality constraints, our method overcomes this limitation and extends to general convex inequality constraints.

Our algorithm also has the ability to achieve exponential convergence and this represents an advancement over existing works that typically ensure only asymptotic or semi-global exponential convergence. On the other hand, our method follows a new line of exponential convergence analysis. Unlike existing methods of proving exponential stability by using the tool of IQC which are valid only when the additional constraints are linear or affine, this paper extends the PAL framework to incorporate general convex inequality constraints and achieves strong convexity-concavity in the primal-dual variables, coming with the capability of dealing with general convex nonlinear constraints.

ACKNOWLEDGEMENT

This work is supported by National Natural Science Foundation of China (62363027, 62373180) and Natural Science Foundation of Jiangxi Province (20224BAB202026).

(Received April 9, 2025)

REFERENCES

-
- [1] B. Abbas and H. Attouch: Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator. *Optimization* *64* (2015), 10, 2223–2252. DOI:10.1080/02331934.2014.971412
 - [2] A. A. Adegbege and M. Y. Kim: Saddle-point convergence of constrained primal-dual dynamics. *IEEE Control Systems Lett.* *5* (2021), 4, 1357–1362. DOI:10.1109/LCSYS.2020.3037876
 - [3] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo: Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Image Process.* *19* (2010), 9, 2345–2356. DOI:10.1109/TIP.2010.2047910

- [4] K. J. Arrow, L. Hurwicz, and H. Uzawa: *Studies in Linear and Non-Linear Programming*. Stanford University Press, 1958.
- [5] P. A. Bansode, V. Chinde, S. R. Wagh, R. Pasumathy, and N. M. Singh: On the exponential stability of projected primal-dual dynamics on a riemannian manifold. *arXiv:1905.04521*, 2019.
- [6] T. Başar and G. J. Olsder: *Dynamic Noncooperative Game Theory*. SIAM, 1998.
- [7] A. Beck and M. Teboulle: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Jo. Imaging Sci.* *2* (2009), 1, 183–202. DOI:10.1137/080716542
- [8] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski: *Robust optimization*. In: *Robust Optimization*, Princeton University Press, 2009.
- [9] M. Benzi, G. H. Golub, and J. Liesen: Numerical solution of saddle point problems. *Acta Numerica* *14* (2005), 1–137. DOI:10.1017/S0962492904000212
- [10] P. J. Bickel and E. Levina: Regularized estimation of large covariance matrices. *Ann. Statist.* *36* (2008), 1, 199–227. DOI:10.1214/009053607000000758
- [11] M. Bin, I. Notarnicola, and T. Parisini: Semiglobal exponential stability of the discrete-time Arrow–Hurwicz–Uzawa primal-dual algorithm for constrained optimization. *Math. Programm.* *208* (2024), 1, 629–660. DOI:10.1007/s10107-023-02051-2
- [12] J. Bolte and M. Teboulle: Barrier operators and associated gradient-like dynamical systems for constrained minimization problems. *SIAM J. Control Optim.* *42* (2003), 4, 1266–1292. DOI:10.1137/S0363012902410861
- [13] S. Boyd and L. Vandenberghe: *Convex Optimization*. Cambridge University Press, 2004. DOI:10.1017/S275390670000070X
- [14] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Machine Learn.* *3* (2011), 1, 1–122. DOI:10.1561/22000000016
- [15] L. M. Bregman: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* *7* (1967), 3, 200–217. DOI:10.1016/0041-5553(67)90040-7
- [16] F. E. Browder: Multi-valued monotone nonlinear mappings and duality mappings in banach spaces. *Trans. Amer. Math. Soc.* *118* (1965), 338–351. DOI:10.1090/S0002-9947-1965-0180884-9
- [17] R. E. Bruck Jr.: On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space. *J. Math. Anal. Appl.* *61* (1977), 1, 159–164. DOI:10.1016/0022-247x(77)90152-4
- [18] C. Charalamous: Nonlinear least pth optimization and nonlinear programing. *Math. Program* *12* (1977), 1, 195–225. DOI:10.1007/bf01593788
- [19] A. Cherukuri, B. Gharesifard, and J. Cortes: Saddle-point dynamics: conditions for asymptotic stability of saddle points. *SIAM J. Control Optim.* *55* (2017), 1, 486–511. DOI:10.1137/15M1026924
- [20] A. Cherukuri, E. Mallada, and J. Cortés: Asymptotic convergence of constrained primal-dual dynamics. *Systems Control Lett.* *87* (2016), 10–15. DOI:10.1016/j.sysconle.2015.10.006
- [21] A. Cherukuri, E. Mallada, S. Low, and J. Cortés: The role of convexity in saddle-point dynamics: Lyapunov function and robustness. *IEEE Trans. Automat. Control* *63* (2018), 8, 2449–2464. DOI:10.1109/TAC.2017.2778689

- [22] P. Cisneros-Velarde, S. Jafarpour, and F. Bullo: Distributed and time-varying primal-dual dynamics via contraction analysis. *arXiv:2003.12665*, 2021.
- [23] P. L. Combettes and V. R. Wajs: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* *4* (2005), 4, 1168–1200. DOI:10.1137/050626090
- [24] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović: The proximal augmented Lagrangian method for nonsmooth composite optimization. *IEEE Trans. Automat. Control* *64* (2019), 7, 2861–2868. DOI:10.1109/TAC.2018.2867589
- [25] N. K. Dhingra, S. Z. Khong, and M. R. Jovanovic: A second order primal-dual method for nonsmooth convex composite optimization. *IEEE Trans. Automat. Control* *67* (2021), 8, 4061–4076. DOI:10.1109/TAC.2021.3115449
- [26] D. Ding and M. R. Jovanovic: Global exponential stability of primal-dual gradient flow dynamics based on the proximal augmented Lagrangian: a Lyapunov-based approach. In: *Proc. 59th IEEE Conference on Decision and Control*, Jeju Island, 2020.
- [27] J. Douglas and H. H. Rachford: On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.* *82* (1956), 2, 421–439. DOI:10.1090/S0002-9947-1956-0084194-4
- [28] J. Eckstein and Ch. Yu: Two innovations in inexact augmented Lagrangian methods for convex optimization. *arXiv preprint arXiv:2503.11809*, 2025.
- [29] M. Fazlyab, A. Ribeiro, M. Morari, and V. M. Preciado: Analysis of optimization algorithms via integral quadratic constraints: Nonstrongly convex problems. *SIAM J. Optim.* *28* (2018), 3, 2654–2689. DOI:10.1137/17M1136845
- [30] D. Feijer and F. Paganini: Stability of primal-dual gradient dynamics and applications to network optimization. *Automatica* *46* (2010), 12, 1974–1981. DOI:10.1016/j.automatica.2010.08.011
- [31] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright: Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Selected Topics Signal Process.* *1* (2007), 4, 586–597. DOI:10.1109/jstsp.2007.910281
- [32] G. França, D. P. Robinson, and R. Vidal: Gradient flows and proximal splitting methods: A unified view on accelerated and stochastic optimization. *Physical Rev. E* *103* (2021), 5, 053304. DOI:10.1103/physreve.103.053304
- [33] D. Goldszajn and F. Paganini: Proximal regularization for the saddle point gradient dynamics. *IEEE Trans. Automat. Control* *66* (2021), 9, 4385–4392. DOI:10.1109/TAC.2020.3045124
- [34] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio: Generative adversarial nets. *Commun. ACM* *63* (2020), 11, 139–144. DOI:10.1145/3422622
- [35] S. Hassan-Moghaddam and M. R. Jovanović: Distributed proximal augmented Lagrangian method for nonsmooth composite optimization. In: *2018 Annual American Control Conference 2018*, pp. 2047–2052.
- [36] S. Hassan-Moghaddam and M. R. Jovanovic: Global exponential stability of the douglas-rachford splitting dynamics. *IFAC-PapersOnLine* *53* (2020), 2, 7350–7354. DOI:10.1016/j.ifacol.2020.12.1254
- [37] S. Hassan-Moghaddam and M. R. Jovanović: Proximal gradient flow and douglas-rachford splitting dynamics: global exponential stability via integral quadratic constraints. *Automatica* *123* (2021), 109311. DOI:10.1016/j.automatica.2020.109311

- [38] M. Hast, K. J. Åström, B. Bernhardsson, and S. Boyd: Pid design by convex-concave optimization. In: 2013 European Control Conference 2013, pp.4460–4465. DOI:10.23919/ecc.2013.6669312
- [39] M. R. Hestenes: Multiplier and gradient methods. *J. Optim. Theory Appl.* *4* (1969), 5, 303–320. DOI:10.1007/BF00927673
- [40] J. B. Hiriart-Urruty and C. Lemaréchal: *Fundamentals of Convex Analysis*. Springer Science and Business Media, 2004.
- [41] X. Ju, H. Che, Ch. Li, and X. He: Solving mixed variational inequalities via a proximal neurodynamic network with applications. *Neural Process. Lett.* *54* (2022), 1, 207–226. DOI:10.1007/s11063-021-10628-1
- [42] H. K. Khalil: *Nonlinear Systems*. Prentice-Hall, Upper Saddle River, NJ 1996.
- [43] T. Kose: Solutions of saddle value problems by differential equations. *Econometrica*, J. Econometr. Soc. 1956, pp. 59–70.
- [44] Y. LeCun, Y. Bengio, and G. Hinton: Deep learning. *Nature* *521* (2015), 7553, 436–444. DOI:10.1038/nature14539
- [45] L. Lessard, B. Recht, and A. Packard: Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM J. Optim.* *26* (2016), 1, 57–95. DOI:10.1137/15M1009597
- [46] F. Lin, M. Fardad, and M. R. Jovanović: Design of optimal sparse feedback gains via the alternating direction method of multipliers. *IEEE Trans. Automat. Control* *58* (2013), 9, 2426–2431. DOI:10.1109/TAC.2013.2257618
- [47] M. Mäkelä: Survey of bundle methods for nonsmooth optimization. *Optim. Methods Software* *17* (2002), 1, 1–29. DOI:10.1080/10556780290027828
- [48] G. J. Minty: Monotone (nonlinear) operators in hilbert space. *Duke Math. J.* *29* (1962), 3, 341–346.
- [49] J.-J. Moreau: Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France* *93* (1965), 273–299.
- [50] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro: Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* *19* (2009), 4, 1574–1609. DOI:10.1137/070704277
- [51] A. S. Nemirovsky and D. B. Yudin: *Problem Complexity and Method Efficiency in Optimization*. John Wiley and Sons, Chichester 1983.
- [52] Y. Nesterov: Smooth minimization of non-smooth functions. *Math. Programm.* *103* (2005), 1, 127–152. DOI:10.1007/s10107-004-0552-5
- [53] J. v Neumann: Zur theorie der gesellschaftsspiele. *Math. Ann.* *100* (1928), 1, 295–320. DOI:10.1007/BF01448847
- [54] J. Nocedal and S. J. Wright: *Numerical Optimization*. Springer, 1999.
- [55] G. B. Passty: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.* *72* (1979), 2, 383–390. DOI:10.1016/0022-247x(79)90234-8
- [56] M. J. D. Powell: A method for nonlinear constraints in minimization problems. *Optimization* (1969), 283–298.
- [57] G. Qu and N. Li: On the exponential stability of primal-dual gradient dynamics. *IEEE Control Systems Lett.* *3* (2019), 1, 43–48. DOI:10.1109/LCSYS.2018.2851375

- [58] M. Raginsky and J. Bouvrie: Continuous-time stochastic mirror descent on a network: variance reduction, consensus, convergence. In: 51st IEEE Conference on Decision and Control 2012, pp. 6793–6800.
- [59] R. T. Rockafellar: Convex Analysis. Princeton University Press, 1970.
- [60] R. T. Rockafellar: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* 1 (1976), 2, 97–116.
- [61] R. T. Rockafellar: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* 14 (1976), 5, 877–898. DOI:10.1137/0314056
- [62] N. Z. Shor: Minimization Methods for Non-differentiable Functions. Springer Science and Business Media, 2012.
- [63] M. Sion: On general minmax theorems. *Pacific J. Math.* 8 (1958), 1, 171–176.
- [64] Y. Tang, G. Qu, and N. Li: Semi-global exponential stability of augmented primal-dual gradient dynamics for constrained convex optimization. *Systems Control Lett.* 144 (2020), 104754. DOI:10.1016/j.sysconle.2020.104754
- [65] R. Tibshirani: Regression shrinkage and selection via the lasso. *J. Royal Statist. Society: Series B (Methodological)* 58 (1996), 1, 267–288. DOI:10.1111/j.2517-6161.1996.tb02080.x
- [66] P. Tseng: A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* 38 (2000), 2, 431–446. DOI:10.1137/S0363012998338806
- [67] G. Wachsmuth: On LICQ and the uniqueness of Lagrange multipliers. *Oper. Res. Lett.* 41 (2013), 1, 78–80. DOI:10.1016/j.orl.2012.11.009
- [68] Z. Wang, W. Wei, Ch. Zhao, Z. Ma, Z. Zheng, Y. Zhang, and F. Liu: Exponential stability of partial primal-dual gradient dynamics with nonsmooth objective functions. *Automatica* 129 (2021), 109585. DOI:10.1016/j.automatica.2021.109585

*Wei Ni, School of Mathematics and Computer Sciences, Nanchang University, Nanchang, P. R. China. He is also with Key Laboratory of Intelligent Systems and Human-Machine Interaction of Jiangxi Province.
e-mail: niwei@ncu.edu.cn*

*Yanyan Xiao, School of Mathematics and Computer Sciences, Nanchang University, Nanchang, P. R. China.
e-mail: 15970167278@163.com*

*Yangfan Qiu, School of Mathematics and Computer Sciences, Nanchang University, Nanchang, P. R. China.
e-mail: 875877601@qq.com*