2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

On the asymptotic stability of proximal algorithms for convex optimization problems with multiple non-smooth regularizers

Ibrahim K. Ozaslan, Sepideh Hassan-Moghaddam, and Mihailo R. Jovanović

Abstract— **We consider composite optimization problems in which the objective function is given by the sum of a smooth convex term and multiple, potentially non-differentiable, convex regularizers. We show that a primal-dual method based on the proximal augmented Lagrangian, which was originally introduced for problems with two components, can be directly extended to this multi-block case. Moreover, we prove that the continuous-time primal-dual dynamics resulting from the proximal augmented Lagrangian are globally asymptotically stable even in the multi-block case if the set of equilibrium points is compact. This is in contrast to ADMM where additional assumptions, e.g., strong convexity of some components, are required. We then examine three-block problems with two non-smooth regularizers and establish global asymptotic stability of splitting dynamic resulting from the proximal augmented Lagrangian.**

Index Terms— **Asymptotic stability, composite optimization, gradient flow dynamics, Lyapunov functions, proximal operator, proximal augmented Lagrangian, operator splitting.**

I. Introduction

We consider a class of composite optimization problems with the objective function given by a sum of a smooth convex term and multiple, possibly non-differentiable, convex regularizers. Such problems arise in a variety of fields including machine learning [1], [2], image processing [3], statistics [4], and control theory [5]–[7].

In multi-block optimization, it is advantageous to introduce auxiliary variables. This decouples non-differentiable parts of the objective function from differentiable ones and facilitates their separate treatment. For problems with two blocks, a successful instance of this approach is given by the celebrated Douglas-Rachford (DR) splitting algorithm [8]. When applied to the Fenchel dual problem, DR splitting gives the Alternating Direction of Method of Multipliers (ADMM) [9], which is widely used in practice [4]. Even though convergence properties of ADMM and DR splitting algorithms are well understood for two-block problems [10]–[14], direct extension of ADMM to three-block problems may not converge without additional assumptions [15]. Hence, extensive efforts have been made to understand additional conditions for ensuring global convergence of ADMM in multi-block case; see [16], [17] and reference therein. In addition to identifying sufficient conditions for convergence, several variants of standard ADMM algorithm that result from different operator splitting rather than DR have also been studied [18], [19]. To the best of our knowledge, unless the smooth part of the objective function is strongly convex or the associated Lagrangian satisfies Kurdyka-Lojasiewicz condition [20] the global convergence of ADMM in multi-block case is an open question.

Several operator splitting schemes including DR and ADMM are particular instances of proximal point iterations [21], [22]. The Method of Multipliers (MM) [4], [23], a technique broadly used for solving constrained nonlinear programming problems, represents another important instance within this class. MM can be derived by applying proximal point iterations to the Fenchel dual problem [24] and, in contrast to ADMM, there are systematic ways for setting algorithmic parameters. However, MM involves joint minimization of the augmented Lagrangian over all primal optimization variables and this task is typically as challenging as solving the original optimization problem. Moreover, joint minimization step in MM precludes parallelization which is indispensable for large-scale problems.

In [25], instead of minimizing the augmented Lagrangian with respect to all primal variables, an approach that constrains it along the manifold resulting from explicit minimization over the variables appearing in non-smooth components was introduced. This approach yields the Proximal Augmented Lagrangian (PAL), which is determined by the sum of original smooth terms and Moreau envelopes associated with non-differentiable regularizers. In contrast to the augmented Lagrangian, PAL is a continuously differentiable function of primal and dual variable and standard techniques from smooth optimization, including Arrow-Hurwicz-Uzawa primal-dual dynamics [26], can be used to compute its saddle points. In the two-block case, these primal-dual dynamics are shown to be globally exponentially stable in both continuous [25] and discrete [27] time. Moreover, methods based on proximal augmented Lagrangian are also convenient for distributed optimization [28] as well as for the development of second-order techniques for non-smooth composite optimization [29].

Studying optimization algorithms as continuous-time dynamical systems has a rich history, starting with the seminal paper by Arrow, Hurwicz, and Uzawa [26]. This viewpoint has recently been advanced and extended to a broad range of problems including convergence analysis of primal-dual [25],

Financial support from the National Science Foundation under awards ECCS-1739210 and ECCS-1809833 is gratefully acknowledged.

I. K. Ozaslan and M. R. Jovanović are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089. E-mails: ozaslan@usc.edu, mihailo@usc.edu.

S. Hassan-Moghaddam is with Microsoft, Redmond, WA 98052.

978-1-6654-5197-0/$31.00 ©2022 AACC

[30]–[34] and accelerated [35]–[41] first-order methods.

In this paper, we study continuous-time primal-dual dynamics based on proximal augmented Lagrangian for multi-block convex composite optimization problems. In addition to smoothness of at least one component in the objective function, we assume that the set of the solutions is compact and prove global asymptotic stability of these dynamics in the multi-block case. We also confine our attention to three-block problems with two non-smooth convex regularizers and show that a splitting dynamic based on PAL represents a continuous-time analogue of the splitting algorithm in [19].

The rest of the paper is organized as follows. In Section II, we formulate a convex multi-block composite optimization problem and provide background material. In Section III-A, we introduce primal-dual gradient flow dynamics based on proximal augmented Lagrangian and employ a Lyapunov-based approach to show that the associated set of equilibrium points is globally asymptotically stable. In Section III-B, we examine a class of three-block problems with two non-smooth regularizers and establish global asymptotic stability of the splitting algorithm resulting from PAL. We conclude the paper in Section IV with remarks.

II. Problem formulation and background

We study composite optimization problems of the form,

$$\underset{x}{\text{minimize}} \;\; f(x) \, + \, \sum_{i\,=\,1}^{r} g_i(T_i x) \tag{1}$$

where $x \in \mathbb{R}^n$ is the optimization variable, f: $\mathbb{R}^n \to \mathbb{R}$ is a convex continuously differentiable function with a Lipschitz continuous gradient ∇f, g_i: $\mathbb{R}^{m_i} \to \mathbb{R}$ are possibly non-differentiable convex regularization functions, and $T_i \in \mathbb{R}^{m_i \times n}$ are given matrices. Equivalently, (1) can be cast as

$$\begin{array}{ll} \underset{x,\,z}{\text{minimize}} & f(x) \, + \, \displaystyle\sum_{i\,=\,1}^{r} g_i(z_i) \\ \text{subject to} & T_i x \, - \, z_i \, = \, 0, \; i \, = \, 1, \ldots, r \end{array} \tag{2}$$

where $z := \left[\begin{array}{ccc} z_1^T & \cdots & z_r^T \end{array}\right]^T$ is the vector of auxiliary variables $z_i \in \mathbb{R}^{m_i}$.

The proximal operator of a proper, closed, convex function g is the minimizer of a quadratically-augmented version of g,

$$\mathbf{prox}_{\mu g}(v) \; := \; \underset{z}{\text{argmin}} \, \left(g(z) \, + \, \frac{1}{2\mu} \, \|z \, - \, v\|_2^2 \right)$$

where μ is a positive parameter and v is a given vector. The value function of this optimization problem determines the associated Moreau envelope,

$$M_{\mu g}(v) \; := \; g(\mathbf{prox}_{\mu g}(v)) \, + \, \frac{1}{2\mu} \, \|\mathbf{prox}_{\mu g}(v) \, - \, v\|_2^2$$

which is a continuously differentiable function, even for a non-differentiable g [21],

$$\mu \nabla M_{\mu g}(v) \; = \; v \, - \, \mathbf{prox}_{\mu g}(v). \tag{3}$$

The Lagrangian associated with (2) is given by,

$$\mathcal{L}(x, z; y) \; := \; f(x) \, + \, \sum_i \left(g_i(z_i) + \langle y_i, T_i x - z_i \rangle \right)$$

and the corresponding augmented Lagrangian is

$$\begin{aligned} \mathcal{L}_\mu(x, z; y) \; &:= \; \mathcal{L}(x, z; y) \, + \, \sum_i \tfrac{1}{2\mu_i} \|T_i x - z_i\|_2^2 \; = \\ f(x) + &\sum_i \left(g_i(z_i) + \tfrac{1}{2\mu_i} \|z_i - (T_i x + \mu_i y_i)\|_2^2 - \tfrac{\mu_i}{2} \|y_i\|_2^2 \right) \end{aligned}$$

where $y := \left[\begin{array}{ccc} y_1^T & \cdots & y_r^T \end{array}\right]^T$ is the vector of Lagrange multipliers $y_i \in \mathbb{R}^{m_i}$ and μ_i's are positive parameters.

The first-order optimality conditions for (2) are

$$0 \; = \; \nabla f(x^\star) \, + \, \sum_i T_i^T y_i^\star \tag{4a}$$

$$0 \; \in \; \partial g_i(z_i^\star) \, - \, y_i^\star, \; i \, = \, 1, \ldots, r \tag{4b}$$

$$0 \; = \; T_i x^\star \, - \, z_i^\star, \; i \, = \, 1, \ldots, r \tag{4c}$$

and these are equivalent to the optimality condition for (1),

$$0 \; \in \; \nabla f(x^\star) \, + \, \sum_i T_i^T \partial g_i(T_i(x^\star)) \tag{5}$$

where ∂g_i is the subdifferntial of g_i. On the other hand, the minimizer of $\mathcal{L}_\mu$ with respect to z_i is determined by

$$z_i^\star(x, y_i) \; = \; \mathbf{prox}_{\mu_i g_i}(T_i x \, + \, \mu_i y_i), \; i \, = \, 1, \ldots, r \tag{6}$$

and the proximal augmented Lagrangian is obtained by evaluating $\mathcal{L}_\mu$ along the manifold determined by $z^\star(x, y)$ [25],

$$\begin{aligned} \mathcal{L}_\mu(x; y) \; &:= \; \mathcal{L}_\mu(x, z^\star(x, y); y) \\ &= \; f(x) + \sum_{i\,=\,1}^{r} \left(M_{\mu_i g_i}(T_i x + \mu_i y_i) - \tfrac{\mu_i}{2} \|y_i\|_2^2 \right). \end{aligned} \tag{7}$$

Here, $M_{\mu_i g_i}$ is the Moreau envelope of the function g_i and the saddle points of (7) are characterized by

$$0 \; = \; \nabla f(x^\star) \, + \, \sum_i T_i^T \nabla M_{\mu_i g_i}(T_i x^\star + \mu_i y_i^\star) \tag{8a}$$

$$0 \; = \; \mu_i \nabla M_{\mu_i g_i}(T_i x^\star + \mu_i y_i^\star) \, - \, \mu_i y_i^\star, \; i \, = \, 1, \ldots, r. \tag{8b}$$

We next show that the saddle points $(x^\star, y^\star)$ of the proximal augmented Lagrangian, determined by (8), satisfy optimality condition (5). Substitution of expression (3) for $\mu_i \nabla M_{\mu_i g_i}$ into (8b) yields

$$T_i x^\star \; = \; \mathbf{prox}_{\mu_i g_i}(T_i x^\star + \mu_i y_i^\star), \; i \, = \, 1, \ldots, r. \tag{9}$$

Since the resolvent operator associated with ∂g_i is single valued and $\mathbf{prox}_{\mu_i g_i} \; = \; (I \, + \, \mu_i \partial g_i)^{-1}$, equation (9) is equivalent to $y_i^\star \in \partial g_i(T_i x^\star)$ for $i = 1, \ldots, r$. Combining this relation with (8b) and (8a) gives (5).

III. Global asymptotic stability of proximal gradient flow algorithms

In this section, we introduce a continuous-time system based on Arrow-Hurwicz-Uzawa (AHU) gradient flow dynamics that can be used to compute saddle points of the proximal augmented Lagrangian (7). For convex problems, we utilize a Lyapunov-based approach to show that these saddle points correspond to the set of globally asymptotically stable equilibrium points of AHU dynamics. Furthermore, for problem (1) with two non-smooth regularizers and $T_1 = T_2 = I$, we utilize a nonlinear coordinate transformation, inspired by Douglas-Rachford splitting [8], to show that the approach based on the proximal augmented Lagrangian yields a continuous-time version of the splitting algorithm developed in [19]. For convex problems, we prove the global asymptotic stability of the set of equilibrium points of the resulting dynamics.

A. Arrow-Hurwicz-Uzawa gradient flow dynamics

The Arrow-Hurwicz-Uzawa gradient flow dynamics based on proximal augmented Lagrangian (7) are given by

$$\begin{aligned} \dot{x} &= -\nabla f(x) - \sum_i T_i^T \nabla M_{\mu_i g_i}(T_i x + \mu_i y_i) \\ \dot{y}_i &= \mu_i \left(\nabla M_{\mu_i g_i}(T_i x + \mu_i y_i) - y_i\right), \; i = 1, \ldots, r. \end{aligned} \tag{10}$$

We next demonstrate that these primal-descent dual-ascent dynamics can be used to compute saddle points of (7) and, thus, to solve convex composite optimization problem (1).

Assumption 1: Let the function f in (1) be convex with an L_f-Lipschitz continuous gradient ∇f and let the regularization functions g_i be proper, closed, and convex.

Assumption 2: The set of solutions to problem (1) characterized by (5) and subdifferentials ∂g_i evaluated at $T_i x^\star$ for each $i = 1, \ldots, r$ are compact.

Remark 1: Assumption 2 provides sufficient conditions for the equilibrium points of gradient flow dynamics (10), which are characterized by (8), to form a compact set. The condition on compactness of the solution set of problem (1) is satisfied if, for example, the function f is strictly convex or if at least one of the regularizers is an indicator function of a compact set. On the other hand, the condition on compactness of $\partial g_i(T_i x^\star)$ is satisfied if $T_i x^\star$ is in the interior domain of g_i [42] or if the functions g_i are locally Lipschitz continuous [43] for all $i = 1, \ldots, r$.

Theorem 1: Let Assumptions 1 and 2 hold. Then, the set of saddle points $(x^\star, y^\star)$ of proximal augmented Lagrangian (7) is a globally asymptotically stable equilibrium set of gradient flow dynamics (10) and each $x^\star$ is a solution of (1).

Proof: Let us introduce a Lyapunov function candidate

$$V(\tilde{x}, \tilde{y}) \;=\; \frac{1}{2}\langle \tilde{x}, \tilde{x}\rangle \;+\; \frac{1}{2}\sum_i \langle \tilde{y}_i, \tilde{y}_i\rangle$$

where $\tilde{x} := x - x^\star$, $\tilde{y}_i := y_i - y_i^\star$, and $(x^\star, y^\star)$ is an equilibrium point of (10) determined by (8). For notational convenience, we define

$$\begin{aligned} \tilde{z}_i &:= z_i^\star(x, y_i) - z_i^\star(x^\star, y_i^\star) \\ &= \mathbf{prox}_{\mu_i g_i}(T_i x + \mu_i y_i) - \mathbf{prox}_{\mu_i g_i}(T_i x^\star + \mu_i y_i^\star) \end{aligned}$$

and express dynamics (10) in new coordinates as

$$\begin{aligned} \dot{\tilde{x}} &= -(\nabla f(x) - \nabla f(x^\star)) - \sum_i \tfrac{1}{\mu_i} T_i^T (T_i \tilde{x} + \mu_i \tilde{y}_i - \tilde{z}_i) \\ \dot{\tilde{y}}_i &= T_i \tilde{x} - \tilde{z}_i, \; i = 1, \ldots, r \end{aligned} \tag{11}$$

by utilizing

$$\nabla M_{\mu_i g_i}(T_i x + \mu_i y_i) - \nabla M_{\mu_i g_i}(T_i x^\star + \mu_i y_i^\star) = \tfrac{1}{\mu_i}(T_i \tilde{x} + \mu_i \tilde{y}_i - \tilde{z}_i).$$

The derivative of V along the solutions of (11) is given by

$$\begin{aligned} \dot{V} &= -\langle \tilde{x}, \nabla f(x) - \nabla f(x^\star)\rangle - \sum_i \tfrac{1}{\mu_i}\|T_i \tilde{x}\|_2^2 + \\ &\qquad \sum_i \tfrac{1}{\mu_i}\langle T_i \tilde{x} - \mu_i \tilde{y}_i, \tilde{z}_i\rangle \\ &= -\langle \tilde{x}, \nabla f(x) - \nabla f(x^\star)\rangle - \sum_i \tfrac{1}{\mu_i}\|T_i \tilde{x}\|_2^2 + \\ &\qquad \sum_i \tfrac{2}{\mu_i}\langle T_i \tilde{x}, \tilde{z}_i\rangle - \sum_i \tfrac{1}{\mu_i}\langle T_i \tilde{x} + \mu_i \tilde{y}_i, \tilde{z}_i\rangle . \end{aligned}$$

Now, since f is convex with an L_f-Lipschitz continuous gradient [44] and $\mathbf{prox}_{\mu_i g_i}$ is firmly non-expansive [21], we have

$$\begin{aligned} -\langle \tilde{x}, \nabla f(x) - \nabla f(x^\star)\rangle &\le -\tfrac{1}{L_f}\|\nabla f(x) - \nabla f(x^\star)\|_2^2 \\ -\langle T_i \tilde{x} + \mu_i \tilde{y}_i, \tilde{z}_i\rangle &\le -\|\tilde{z}_i\|_2^2 \end{aligned} \tag{12}$$

and substitution of these upper bounds to $\dot{V}$ yields

$$\dot{V} \;\le\; -\tfrac{1}{L_f}\|\nabla f(x) - \nabla f(x^\star)\|_2^2 - \sum_i \tfrac{1}{\mu_i}\|T_i \tilde{x} - \tilde{z}_i\|_2^2.$$

Thus, $\dot{V}$ is a negative semidefinite function and the set of equilibrium points is stable in the sense of Lyapunov.

Furthermore, since $\dot{V} \le 0$ for the points satisfying $\nabla f(x) = \nabla f(x^\star)$ and $T_i \tilde{x} = \tilde{z}_i$, dynamics (11) evaluated at these points become

$$\begin{aligned} \dot{\tilde{x}} &= -\sum_i T_i^T \tilde{y}_i \\ \dot{\tilde{y}}_i &= 0, \; i = 1, \ldots, r \end{aligned}$$

and $\dot{V} = \langle \tilde{x}, \sum_i T_i^T \tilde{y}_i\rangle$. Now, let

$$\mathcal{C} := \{(x, y) \,|\, \dot{V}(\tilde{x}, \tilde{y}) = 0\}$$

and note that if $\langle \tilde{x}, \sum_i T_i^T \tilde{y}_i\rangle = 0$, then

$$\begin{aligned} 0 &= \tfrac{\mathrm{d}}{\mathrm{d}t}\langle \tilde{x}, \textstyle\sum_i T_i^T \tilde{y}_i\rangle \\ &= \langle \dot{\tilde{x}}, \textstyle\sum_i T_i^T \tilde{y}_i\rangle + \langle \tilde{x}, \textstyle\sum_i T_i^T \dot{\tilde{y}}_i\rangle \\ &= \langle \dot{\tilde{x}}, \textstyle\sum_i T_i^T \tilde{y}_i\rangle \\ &= -\|\textstyle\sum_i T_i^T \tilde{y}_i\|_2^2. \end{aligned}$$

This implies that $\sum_i T_i^T \tilde{y}_i = 0$. Thus, under dynamics (11), the largest invariant set $\Omega \subseteq \mathcal{C}$ is given by

$$\Omega := \{(\tilde{x}, \tilde{y}) \,|\, (\tilde{x}, \tilde{y}) \in \mathcal{C}, \; \textstyle\sum_i T_i^T \tilde{y}_i = 0\},$$

Next, we show that Ω is equivalent to the set of equilibrium points characterized by (8). For every pair (x, y) such that

$(\tilde{x}, \tilde{y}) \in \Omega$, we have

$$\nabla f(x) - \nabla f(x^\star) = 0 \quad (13a)$$

$$\sum_i T_i^T (y_i - y_i^\star) = 0 \quad (13b)$$

$$T_i x - \mathbf{prox}_{\mu_i g_i}(T_i x + \mu_i y_i) = 0, \; i = 1, \ldots, r. \quad (13c)$$

Equality (13c) follows from $T_i \tilde{x} = \tilde{z}_i$ and (9), and is equivalent to $y_i = \nabla M_{\mu_i g_i}(T_i x + \mu_i y)$ for $i = 1, \ldots, r$. Substituting these relations to (13b) and then adding (13b) to (13a) yield optimality condition (8a). Finally, (8b) can be obtained from (13c) using (3). This shows that Ω is a compact set and LaSalle's Invariance Principle implies global asymptotic stability of the set of equilibrium points. ∎

B. A splitting algorithm for problems with two regularizers

For two regularization functions with $T_1 = T_2 = I$ in (1),

$$\underset{x}{\text{minimize}} \; f(x) + g_1(x) + g_2(x) \quad (14)$$

optimality conditions (8) for proximal augmented Lagrangian (7) with $\mu_1 = \mu_2 = \mu$ simplify to

$$\begin{aligned} \nabla f(x^\star) + \nabla M_{\mu g_1}(x^\star + \mu y_1^\star) + \nabla M_{\mu g_2}(x^\star + \mu y_2^\star) &= 0 \\ \nabla M_{\mu g_1}(x^\star + \mu y_1^\star) - y_1^\star &= 0 \\ \nabla M_{\mu g_2}(x^\star + \mu y_2^\star) - y_2^\star &= 0. \end{aligned}$$

The last two equalities together with (3) yield

$$x^\star = \mathbf{prox}_{\mu g_1}(x^\star + \mu y_1^\star) = \mathbf{prox}_{\mu g_2}(x^\star + \mu y_2^\star) \quad (15)$$

and from the above optimality conditions we also have

$$\nabla f(x^\star) + y_1^\star + y_2^\star = 0. \quad (16)$$

Now, introduction of an auxiliary variable

$$\xi^\star := x^\star + \mu y_1^\star \quad (17a)$$

allows us to express $x^\star$ as

$$x^\star = \mathbf{prox}_{\mu g_1}(\xi^\star) \quad (17b)$$

and combine (16) with (17a) and (17b) to obtain

$$\begin{aligned} x^\star + \mu y_2^\star &= x^\star - \mu \nabla f(x^\star) - \mu y_1^\star \\ &= 2x^\star - \mu \nabla f(x^\star) - \xi^\star \\ &= 2\mathbf{prox}_{\mu g_1}(\xi^\star) - \mu \nabla f(\mathbf{prox}_{\mu g_1}(\xi^\star)) - \xi^\star. \end{aligned} \quad (17c)$$

Finally, substitution of (17a) and (17c) into (15) yields the following optimality condition,

$$P_{\mu g_1}(\xi^\star) = P_{\mu g_2}(2P_{\mu g_1}(\xi^\star) - \mu \nabla f(P_{\mu g_1}(\xi^\star)) - \xi^\star) \quad (18)$$

where, for notational compactness, $P_{\mu g_i} := \mathbf{prox}_{\mu g_i}$ and $x^\star$ is determined by (17b).

We next demonstrate that the proximal splitting dynamics obtained from (18),

$$\dot{\xi} = -P_{\mu g_1}(\xi) + P_{\mu g_2}(2P_{\mu g_1}(\xi) - \mu \nabla f(P_{\mu g_1}(\xi)) - \xi), \quad (19)$$

can be used to solve composite optimization problem (14). We note that the dynamical system in (19) is a continuous-time version of the splitting algorithm developed in [19].

Theorem 2: Let Assumption 1 hold and let $\mu \in (0, 2/L_f)$. Then, the set of equilibrium points $\{z^\star\}$ of proximal splitting dynamics (19) is globally asymptotically stable and corresponding $x^\star = \mathbf{prox}_{\mu g_1}(\xi^\star)$ is an optimal solution of (14).

Proof: Let us introduce a change of variables and a Lyapunov function candidate

$$V(\tilde{\xi}) = \frac{1}{2}\|\tilde{\xi}\|_2^2 = \frac{1}{2}\|\xi - \xi^\star\|_2^2 \quad (20)$$

where $\xi^\star$ satisfies (18). For notational convenience, we also introduce three nonlinear functions,

$$\begin{aligned} \Delta_1(\xi) &:= P_{\mu g_1}(\xi) - \mu \nabla f(P_{\mu g_1}(\xi)) \\ \Delta_2(\xi) &:= P_{\mu g_1}(\xi) + \Delta_1(\xi) \\ \Delta(\xi) &:= \Delta_2(\xi) - \xi \end{aligned}$$

and rewrite (19) in new coordinates,

$$\begin{aligned} \dot{\tilde{\xi}} = & -(P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star)) + \\ & (P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star))). \end{aligned} \quad (21)$$

The derivative of V along the solutions of (21) is given by

$$\begin{aligned} \dot{V} = & -\langle \xi - \xi^\star, P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star) \rangle + \\ & \langle \xi - \xi^\star, P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star)) \rangle. \end{aligned}$$

Since the proximal operator is firmly non-expansive [21], the first term on the right-hand-side satisfies

$$-\langle \xi - \xi^\star, P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star) \rangle \le -\|P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star)\|_2^2$$

and, since $\xi = \Delta_2(\xi) - \Delta(\xi) = P_{\mu g_1}(\xi) + \Delta_1(\xi) - \Delta(\xi)$, the second can be written as

$$\begin{aligned} & \langle \xi - \xi^\star, P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star)) \rangle = \\ & -\langle \Delta(\xi) - \Delta(\xi^\star), P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star)) \rangle + \\ & \langle P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star), P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star)) \rangle + \\ & \langle \Delta_1(\xi) - \Delta_1(\xi^\star), P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star)) \rangle. \end{aligned} \quad (22)$$

For the first term on the right-hand-side of (22) we have the following upper bound,

$$\begin{aligned} & -\langle \Delta(\xi) - \Delta(\xi^\star), P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star)) \rangle \le \\ & -\|P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star))\|_2^2 \end{aligned}$$

and for the third, we use the Fenchel-Young inequality

$$\begin{aligned} & \langle \Delta_1(\xi) - \Delta_1(\xi^\star), P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star)) \rangle \le \\ & \frac{1}{2}\big(\|\Delta_1(\xi) - \Delta_1(\xi^\star)\|_2^2 + \|P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star))\|_2^2\big). \end{aligned}$$

Furthermore, from the definition of $\Delta_1(\xi)$, we have

$$\begin{aligned} & \|\Delta_1(\xi) - \Delta_1(\xi^\star)\|_2^2 = \|P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star)\|_2^2 + \\ & \mu^2 \|\nabla f(P_{\mu g_1}(\xi)) - \nabla f(P_{\mu g_1}(\xi^\star))\|_2^2 - \\ & 2\mu \langle P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star), \nabla f(P_{\mu g_1}(\xi)) - \nabla f(P_{\mu g_1}(\xi^\star)) \rangle. \end{aligned} \quad (23)$$

Since f satisfies (12), we have the following upper bound for (23),

$$\|\Delta_1(\xi) - \Delta_1(\xi^\star)\|_2^2 \;\le\; \|P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star)\|_2^2 \;- \\ \mu(2/L_f - \mu)\|\nabla f(P_{\mu g_1}(\xi)) - \nabla f(P_{\mu g_1}(\xi^\star))\|_2^2. \tag{24}$$

For $\mu \in (0, 2/L_f)$, we can use the above expressions to upper bound the derivative of V along the solutions of (21) with

$$-\tfrac{1}{2}\,\|(P_{\mu g_1}(\xi) - P_{\mu g_1}(\xi^\star)) - (P_{\mu g_2}(\Delta(\xi)) - P_{\mu g_2}(\Delta(\xi^\star))\|_2^2$$

Finally, using (18) we can write

$$\dot{V} \;\le\; -\tfrac{1}{2}\,\|P_{\mu g_1}(\xi) - P_{\mu g_2}(\Delta(\xi))\|_2^2$$

which shows that $\dot{V}$ is a negative definite function everywhere apart from the set of equilibrium points, where $\dot{V} = 0$. This completes the proof. ■

IV. Concluding Remarks

We have considered a class of composite optimization problems where the objective function can be expressed as a sum of a smooth convex term and multiple possibly non-differentiable convex regularizers. We proved that if the set of equilibrium points is compact, then, unlike existing ADMM variants, the Arrow-Hurwicz-Uzawa gradient flow dynamics based on proximal augmented Lagrangian are globally asymptotically stable even in the multi-block case. We then confined our attention to a three-block problem and showed that the splitting dynamics resulting from the proximal augmented Lagrangian represent a continuous-time version of the algorithm developed in [19]. In our ongoing effort, we aim to identify conditions for global exponential stability of the primal-dual gradient flow dynamics based on proximal augmented Lagrangian.

Acknowledgments

We would like to thank Panos Patrinos and Puya Latafat for useful discussions and anonymous reviewers for their insightful comments.

References

[1] M. Pilanci and T. Ergen, “Neural networks are convex regularizers: Exact polynomial-time convex optimization formulations for two-layer networks,” in *International Conference on Machine Learning*. PMLR, 2020, pp. 7695–7705.

[2] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, “Simultaneously structured models with application to sparse and low-rank matrices,” *IEEE Trans. Inform. Theory*, vol. 61, no. 5, pp. 2886–2908, 2015.

[3] V. Cevher, S. Becker, and M. Schmidt, “Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics,” *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 32–43, 2014.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundation and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–124, 2011.

[5] F. Lin, M. Fardad, and M. R. Jovanović, “Augmented Lagrangian approach to design of structured optimal state feedback gains,” *IEEE Trans. Automat. Control*, vol. 56, no. 12, pp. 2923–2929, 2011.

[6] 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*, vol. 58, no. 9, pp. 2426–2431, 2013.

[7] M. R. Jovanović and N. K. Dhingra, “Controller architectures: tradeoffs between performance and structure,” *Eur. J. Control*, vol. 30, pp. 76–91, 2016.

[8] J. Douglas and H. Rachford, “On the numerical solution of heat conduction problems in two and three space variables,” *Trans. Amer. Math. Soc.*, vol. 82, no. 2, pp. 421–439, 1956.

[9] D. Gabay, “Applications of the method of multipliers to variational inequalities,” in *Studies in Mathematics and its Applications*, 1983, vol. 15, pp. 299–331.

[10] P. Patrinos, L. Stella, and A. Bemporad, “Forward-backward truncated Newton methods for convex composite optimization,” 2014, arXiv:1402.6655.

[11] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, “A general analysis of the convergence of ADMM,” in *International Conference on Machine Learning*. PMLR, 2015, pp. 343–352.

[12] P. Giselsson and S. Boyd, “Linear convergence and metric selection for Douglas-Rachford splitting and ADMM,” *IEEE Trans. Automat. Control*, vol. 62, no. 2, pp. 532–544, 2017.

[13] A. Themelis and P. Patrinos, “Douglas-Rachford splitting and ADMM for nonconvex optimization: Tight convergence results,” *SIAM J. Optim.*, vol. 30, no. 1, pp. 149–181, 2020.

[14] S. Hassan-Moghaddam and M. R. Jovanović, “Proximal gradient flow and Douglas-Rachford splitting dynamics: global exponential stability via integral quadratic constraints,” *Automatica*, vol. 123, p. 109311, 2021.

[15] C. Chen, B. He, Y. Ye, and X. Yuan, “The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent,” *Math. Program.*, vol. 155, no. 1-2, pp. 57–79, 2016.

[16] T. Lin, S. Ma, and S. Zhang, “On the global linear convergence of the ADMM with multiblock variables,” *SIAM J. Optim.*, vol. 25, no. 3, pp. 1478–1497, 2015.

[17] T. Lin, S. Ma, and S. Zhang, “Iteration complexity analysis of multi-block ADMM for a family of convex minimization without strong convexity,” *J. Sci. Comput.*, vol. 69, no. 1, pp. 52–81, 2016.

[18] P. Latafat and P. Patrinos, “Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators,” *Comput. Optim. Appl.*, vol. 68, no. 1, pp. 57–93, 2017.

[19] D. Davis and W. Yin, “A three-operator splitting scheme and its optimization applications,” *Set-Valued Var. Anal.*, vol. 25, no. 4, pp. 829–858, 2017.

[20] H. Karimi, J. Nutini, and M. Schmidt, “Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2016, pp. 795–811.

[21] N. Parikh and S. Boyd, “Proximal algorithms,” *Found. Trends Opt.*, vol. 1, no. 3, pp. 123–231, 2013.

[22] J. Eckstein and D. P. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Math. Program.*, vol. 55, no. 1-3, pp. 293–318, 1992.

[23] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic Press, 1982.

[24] E. K. Ryu and S. Boyd, “Primer on monotone operator methods,” *Appl. Comput. Math*, vol. 15, no. 1, pp. 3–43, 2016.

[25] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, “The proximal augmented Lagrangian method for nonsmooth composite optimization,” *IEEE Trans. Automat. Control*, vol. 64, no. 7, pp. 2861–2868, 2019.

[26] K. J. Arrow, L. Hurwicz, and H. Uzawa, “Studies in linear and nonlinear programming,” 1958.

[27] D. Ding, B. Hu, N. K. Dhingra, and M. R. Jovanović, “An exponentially convergent primal-dual algorithm for nonsmooth composite minimization,” in *Proceedings of the 57th IEEE Conference on Decision and Control*, Miami, FL, 2018, pp. 4927–4932.

[28] S. Hassan-Moghaddam and M. R. Jovanović, “Distributed proximal augmented Lagrangian method for nonsmooth composite optimization,” in *Proceedings of the 2018 American Control Conference*, Milwaukee, WI, 2018, pp. 2047–2052.

[29] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović, “A second order primal-dual method for nonsmooth convex composite optimization,” *IEEE Trans. Automat. Control*, 2022, doi:10.1109/TAC.2021.3115449.

[30] D. Feijer and F. Paganini, “Stability of primal–dual gradient dynamics and applications to network optimization,” *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.

[31] J. Wang and N. Elia, “A control perspective for centralized and distributed convex optimization,” in *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011, pp. 3800–3805.

[32] A. Cherukuri, E. Mallada, and J. Cortés, "Asymptotic convergence of constrained primal–dual dynamics," *Syst. Control Lett.*, vol. 87, pp. 10–15, 2016.

[33] A. Cherukuri, E. Mallada, S. Low, and J. Cortes, "The role of convexity on saddle-point dynamics: Lyapunov function and robustness," *IEEE Trans. Automat. Control*, vol. 63, no. 8, pp. 2449–2464, 2018.

[34] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 43–48, 2018.

[35] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights," *J. Mach. Learn. Res.*, vol. 17, pp. 1–43, 2016.

[36] A. Wibisono, A. C. Wilson, and M. I. Jordan, "A variational perspective on accelerated methods in optimization," *Proc. Natl. Acad. Sci.*, vol. 113, no. 47, pp. E7351–E7358, 2016.

[37] G. França, D. Robinson, and R. Vidal, "ADMM and accelerated ADMM as continuous dynamical systems," in *Proceedings of the 35th International Conference on Machine Learning*, 2018, pp. 1559–1567.

[38] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su, "Understanding the acceleration phenomenon via high-resolution differential equations," *Math. Program.*, pp. 1–70, 2021.

[39] M. Muehlebach and M. Jordan, "A dynamical systems perspective on Nesterov acceleration," in *Proceedings of the 36th International Conference on Machine Learning*, 2019, pp. 4656–4662.

[40] J. I. Poveda and N. Li, "Inducing uniform asymptotic stability in non-autonomous accelerated optimization dynamics via hybrid regularization," in *Proceedings of the 58th IEEE Conference on Decision and Control*, 2019, pp. 3000–3005.

[41] G. França, D. P. Robinson, and R. Vidal, "Gradient flows and proximal splitting methods: A unified view on accelerated and stochastic optimization," *Phys. Rev. E*, vol. 103, no. 5, p. 053304, 2021.

[42] H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011. [Online]. Available: https://hal.inria.fr/hal-00643354

[43] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *J. Nonlinear Sci.*, vol. 29, no. 4, pp. 1247–1272, 2019.

[44] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, 2013, vol. 87.