

Accelerated forward-backward and Douglas-Rachford splitting dynamics

Ibrahim K. Ozaslan and Mihailo R. Jovanović

Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089

Abstract

We examine convergence properties of continuous-time variants of accelerated Forward-Backward (FB) and Douglas-Rachford (DR) splitting algorithms for nonsmooth composite optimization problems. When the objective function is given by the sum of a quadratic and a nonsmooth term, we establish accelerated sublinear and exponential convergence rates for convex and strongly convex problems, respectively. Moreover, for FB splitting dynamics, we demonstrate that accelerated exponential convergence rate carries over to general strongly convex problems. In our Lyapunov-based analysis we exploit the variable-metric gradient interpretations of FB and DR splittings to obtain smooth Lyapunov functions that allow us to establish accelerated convergence rates. We provide computational experiments to demonstrate the merits and the effectiveness of our analysis.

Key words: Nesterov’s acceleration; proximal gradient method; forward-backward splitting; Douglas-Rachford splitting; envelope functions; primal-dual gradient flow dynamics.

1 Introduction

The convergence properties of gradient descent (GD),

$$x^{k+1} = x^k - \bar{\alpha}\nabla f(x^k) \quad (1)$$

can be improved by adding a momentum term,

$$x^{k+1} = x^k + \bar{\gamma}(x^k - x^{k-1}) + \bar{\alpha}\nabla f(x^k + \bar{\beta}(x^k - x^{k-1})), \quad (2)$$

where k is the iteration index, $\bar{\alpha}$ is the stepsize, $\bar{\beta}$ is the extrapolation parameter, and $\bar{\gamma}$ is the damping coefficient. In particular, the Polyak’s Heavy-Ball Method (HBM) is obtained by setting $\bar{\beta} = 0$ in (2) [22] and Nesterov’s Accelerated Algorithm (NAA) results from $\bar{\gamma} = \bar{\beta}$ in (2) [14]. For a convex objective function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ with an L -Lipschitz continuous gradient ∇f , NAA with $(\bar{\alpha}, \bar{\beta}) = (1/L, (k-2)/(k+3))$ improves the sublinear convergence rate $O(1/k)$ of GD to $O(1/k^2)$. Furthermore, for an m -strongly convex f , the linear convergence rate $O(e^{-k/\kappa})$ of GD [15, Section 2.1.5] is improved to $O(e^{-k/\sqrt{\kappa}})$ with $(\bar{\alpha}, \bar{\beta}) = (1/L, (\sqrt{\kappa}-1)/(\sqrt{\kappa}+1))$. In both cases, these rates are optimal in the sense that there exists a smooth convex function f for which the rates cannot be improved [15, Theorems 2.1.7 and 2.1.13].

A wide-spread use of accelerated algorithms has inspired efforts to understand the underlying mechanisms. In particular, a growing body of literature views optimization

algorithms as continuous-time dynamical systems that allow familiar methods to be recovered upon proper discretization [1, 6, 7, 13]. For example, the continuous-time equivalent of GD is given by the gradient flow dynamics,

$$\dot{x} = -\alpha\nabla f(x) \quad (3)$$

and celebrated Arrow-Hurwicz-Uzawa primal-dual gradient flow dynamics can be used to effectively solve a class of constrained optimization problems [17]. A dynamical systems perspective of NAA offered in [24] was further exploited by leveraging a variational formulation based on Lagrangian and Hamiltonian frameworks from classical mechanics [25, 26] and contraction theory [5]. Furthermore, NAA was obtained using the semi-implicit Euler discretization of the differential equation [13],

$$\ddot{x} + \gamma\dot{x} + \alpha\nabla f(x + \beta\dot{x}) = 0 \quad (4)$$

where α , β , and γ are possibly time-varying algorithmic parameters that differ for convex and strongly convex problems. The following Lyapunov function,

$$V(x, \dot{x}) = af(x) + (1/2)\|bx + \dot{x}\|_2^2 \quad (5)$$

with positive parameters a and b was also utilized in [13] to show that (4) achieves the optimal convergence rate of discrete-time algorithms for both convex and strongly convex problems with proper choices of (α, β, γ) . Moreover, for $\beta + \gamma = 1$, the discretization of (4) preserves the accelerated convergence rate. This provides a system-theoretic characterization of acceler-

Email addresses: ozaslan@usc.edu (Ibrahim K. Ozaslan and), mihailo@usc.edu (Mihailo R. Jovanović).

ation and helps bridging the gap between discrete and continuous-time analyses.

Our contributions. We generalize accelerated dynamics (4) to nonsmooth composite optimization problems in which the objective function is given by the sum of a smooth and a nonsmooth term. Our approach is motivated by [10] which obtains the Forward-Backward (FB) and Douglas-Rachford (DR) splitting dynamics by replacing ∇f in (3) with the generalized gradient map associated with the nonsmooth objective function. Based on the relation between gradient flow dynamics (3) and its accelerated variant (4), we introduce accelerated FB and DR splitting dynamics by replacing gradient ∇f in (4) with the generalized gradient map. We also utilize a Lyapunov-based approach to establish accelerated sublinear and linear convergence rates that match the optimal discrete-time rates for convex and strongly convex problems, respectively. Our Lyapunov function is similar to (5) with two important differences: (i) the Euclidean distance is replaced with a suitable weighted norm; and (ii) the FB or DR envelopes are used instead of the objective function. The latter choice is motivated by the variable-metric gradient method interpretation of associated splittings [20, 21]. We also note that, in our analysis, the condition $\beta + \gamma = 1$ is satisfied, which is an integral requirement for discretization analysis of [13] to be applicable to the proposed accelerated dynamics.

Although accelerated rates for the FB and DR splittings are known in discrete-time, the continuous-time analysis in the problem settings that we examine has not been done before. An accelerated sublinear convergence rate matching our result has been proposed in [24], but the approach relies on a dynamical system that is based on directional subgradients whose evaluation requires solving linear program $\operatorname{argmax}_{z \in \partial g(x)} z^T \dot{x}$ at every time instant, where g denotes the nonsmooth term in the objective function. Moreover, the existence and uniqueness of a solution to this dynamical system has been assumed without formally establishing it [24, Theorem 24]. An accelerated sublinear convergence rate has also been proposed in [2, 4] for a differential inclusion that models accelerated FB splitting, but no analysis was provided.

It is also worth mentioning that the FB and DR splittings are special cases of the primal-dual gradient flow dynamics associated with the proximal augmented Lagrangian [6, 10, 16]. Thus, insights about acceleration of these splitting methods can also pave the way to accelerated primal-dual algorithms, which is of critical importance to the design of fast gradient-based methods for constrained optimization problems. We note that (4) was used as a motivation in [3, 11, 27] to accelerate primal-dual gradient flow dynamics. However, in contrast to [13], the algorithmic parameters are not bounded in time in these references. Thus, it is unclear if the primal-dual dynamics proposed in [3, 11, 27] allow rate-preserving discretization that leads to efficient

first-order iterative schemes.

Paper structure. In Section 2, we introduce the problem setup and provide background material. In Section 3, we summarize our main findings, and in Section 4 we provide the proofs. In Section 5, we utilize computational experiments to demonstrate the merits of our analyses, and in Section 6, we conclude the paper with remarks.

2 Problem formulation and preliminaries

In this section, we formulate the problem and provide background on the proximal operators; Moreau, FB, and DR envelopes; as well as FB and DR splittings.

We consider convex optimization problem

$$\underset{x}{\operatorname{minimize}} F(x) := f(x) + g(x) \quad (6)$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is, unless noted otherwise, a continuously differentiable convex function with an L -Lipschitz continuous gradient ∇f and $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is a possibly non-differentiable closed proper convex function.

2.1 Proximal operator and Moreau envelope

The proximal operator of a proper lower semi-continuous convex function g for a positive penalty parameter μ is the mapping defined as $\mathbf{prox}_{\mu g}(v) = (I + \mu \partial g)^{-1}(v)$, where ∂g is a subdifferential of g and I is the identity map [19]. Alternatively, $\mathbf{prox}_{\mu g}$ can be obtained as

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

The value function of this optimization problem determines the Moreau envelope associated with g ,

$$\mathcal{M}_{\mu g}(v) = g(\mathbf{prox}_{\mu g}(v)) + \frac{1}{2\mu} \|\mathbf{prox}_{\mu g}(v) - v\|_2^2 \quad (7)$$

which is, even for a non-differentiable g , a continuously differentiable function [19],

$$\nabla \mathcal{M}_{\mu g}(v) = \frac{1}{\mu} (v - \mathbf{prox}_{\mu g}(v)) \in \partial g(\mathbf{prox}_{\mu g}(v)). \quad (8)$$

2.2 The FB and DR envelopes

Using Lipschitz continuity of ∇f , we have the following upper bound on the objective function F in (6) [21],

$$F(y) \leq J(x, y) := f(x) + g(y) + \frac{1}{2\mu} \|y - (x - \mu \nabla f(x))\|_2^2 - \frac{\mu}{2} \|\nabla f(x)\|_2^2$$

which holds for any $x, y \in \mathbb{R}^n$ and $\mu \in (0, 1/L)$. Minimizing $J(x, y)$ over y yields the FB envelope of F [21]

$$\begin{aligned} \mathcal{F}_{\mu}(x) &:= \underset{y}{\operatorname{minimize}} J(x, y) \\ &= f(x) + \mathcal{M}_{\mu g}(x - \mu \nabla f(x)) - \frac{\mu}{2} \|\nabla f(x)\|_2^2. \end{aligned}$$

For a particular choice of a distance function (instead of the Euclidean norm), the FB envelope can also be interpreted as a generalized Moreau envelope of F [12]. The DR envelope associated with (6) is another useful value function resulting from evaluation of the FB envelope $\mathcal{F}_\mu(x)$ at $x = \mathbf{prox}_{\mu f}(z)$ with $z \in \mathbb{R}^n$ [8, 20], i.e.,

$$\mathcal{D}_\mu(z) := \mathcal{F}_\mu(\mathbf{prox}_{\mu f}(z)).$$

When f is a twice continuously differentiable function with a bounded Hessian $\nabla^2 f(x)$, the FB envelope is also continuously differentiable with gradient

$$\nabla \mathcal{F}_\mu(x) = (I - \mu \nabla^2 f(x)) G_\mu(x) \quad (9)$$

where G_μ is the generalized gradient map defined as

$$G_\mu(x) = \frac{1}{\mu}(x - \mathbf{prox}_{\mu g}(x - \mu \nabla f(x))). \quad (10)$$

Consequently, the DR envelope is also continuously differentiable and its gradient is determined by

$$\nabla \mathcal{D}_\mu(z) = \nabla \mathbf{prox}_{\mu f}(z) \nabla \mathcal{F}_\mu(\mathbf{prox}_{\mu f}(z)). \quad (11)$$

Here, $\nabla \mathbf{prox}_{\mu f}(z)$ can be obtained by differentiating (8)

$$\nabla \mathbf{prox}_{\mu f}(z) := (I + \mu \nabla^2 f(\mathbf{prox}_{\mu f}(z)))^{-1} \quad (12)$$

Using (9) and (11), $\nabla \mathcal{D}_\mu$ can also be expressed in terms of the generalized gradient map as

$$\nabla \mathcal{D}_\mu(z) = (2 \nabla \mathbf{prox}_{\mu f}(z) - I) G_\mu(\mathbf{prox}_{\mu f}(z)). \quad (13)$$

We note that for $\mu \in (0, 1/L)$,

$$\begin{aligned} \min F &= \min \mathcal{F}_\mu = \min \mathcal{D}_\mu \\ \operatorname{argmin} F &= \operatorname{argmin} \mathcal{F}_\mu = \mathbf{prox}_{\mu f}(\operatorname{argmin} \mathcal{D}_\mu) \end{aligned} \quad (14)$$

which together with the smoothness of the envelope functions motivates studying acceleration of dynamics (16) and (20) through the FB and DR envelopes [21].

2.3 The FB and DR splittings

Proximal gradient algorithm [19],

$$x^{k+1} = x^k - \bar{\alpha} G_\mu(x^k) \quad (15)$$

replaces ∇f in (1) with the generalized gradient map G_μ , thereby extending gradient descent to composite optimization problem (6). Iterations (15) are also known as the FB splitting and their continuous-time equivalent is determined by the proximal gradient flow dynamics [10],

$$\dot{x} = -\alpha G_\mu(x). \quad (16)$$

Furthermore, evaluation of the generalized gradient map at $x = \mathbf{prox}_{\mu f}(z)$ facilitates extension of FB splitting (15) to composite optimization problem (6) in which both f and g are allowed to be non-differentiable,

$$z^{k+1} \in z^k - \bar{\alpha} G_\mu(\mathbf{prox}_{\mu f}(z^k)). \quad (17)$$

Using (10) and setting $\bar{\alpha} = \mu$ in (17) yields,

$$\begin{aligned} z^{k+1} \in & z^k - \mathbf{prox}_{\mu f}(z) + \\ & \mathbf{prox}_{\mu g}(\mathbf{prox}_{\mu f}(z) - \mu \partial f(\mathbf{prox}_{\mu f}(z))). \end{aligned} \quad (18)$$

Moreover, (8) can be used to rewrite (18) as,

$$z^{k+1} = z^k - \mathbf{prox}_{\mu f}(z) + \mathbf{prox}_{\mu g}(2 \mathbf{prox}_{\mu f}(z) - z) \quad (19)$$

which is known as DR splitting [23]. While (6) can be solved via (19) even for a non-differentiable f , when f is continuously differentiable $G_\mu(\mathbf{prox}_{\mu f}(z))$ becomes a single-valued function and (17) can be used to obtain the DR splitting dynamics [10],

$$\dot{z} = -\alpha G_\mu(\mathbf{prox}_{\mu f}(z)). \quad (20)$$

Lemma 1 characterizes stability properties of proximal gradient flow (16) and DR splitting dynamics (20) under the following assumption.

Assumption 1 *In (6), f is an m -strongly convex function with an L -Lipschitz continuous gradient ∇f and g is a proper closed convex function.*

Lemma 1 (Thms. 3 and 7, [10]) *Let Assumption 1 hold with $m > 0$ and let $\mu = 1/(2L)$. Then, the proximal gradient flow (16) and DR splitting dynamics (20) are globally exponentially stable with rate $\rho = \alpha m$.*

3 Main results

In this section, we examine convergence properties of continuous-time variants of accelerated proximal gradient and DR splitting algorithms for nonsmooth composite optimization problem (6). In Theorems 1 and 2, we show that the corresponding dynamical systems provide acceleration when the smooth part of the objective function f is either convex or strongly convex quadratic function, respectively. Furthermore, in Theorem 3, we demonstrate that properties of accelerated proximal gradient flow dynamics carry over from quadratic to general strongly convex problems. To improve readability, we relegate the proofs of theorems to Sections 4.

Inspired by (4), which provides accelerated convergence for smooth problems, we propose the following differential equation to accelerate proximal gradient flow dynamics (16),

$$\ddot{x} + \gamma \dot{x} + \alpha G_\mu(x + \beta \dot{x}) = 0. \quad (21a)$$

Similarly, we introduce

$$\begin{aligned} \ddot{z} + \gamma\dot{z} + \alpha G_\mu(\mathbf{prox}_{\mu f}(z + \beta\dot{z})) &= 0 \\ x &= \mathbf{prox}_{\mu f}(z) \end{aligned} \quad (21b)$$

as an accelerated variant of DR splitting dynamics (20), where G_μ is the generalized gradient map associated with problem (6); see (10) for definition.

Our analysis of (21a) and (21b) utilizes Lyapunov functions obtained by replacing f in (5) with smooth envelope functions \mathcal{F}_μ or \mathcal{D}_μ , respectively. This choice is motivated by the variable-metric gradient method interpretations of FB and DR splittings. The lack of convexity of the FB and DR envelopes when f in (6) is not a convex quadratic function is the main challenge in convergence analysis. When f is either a convex or a strongly convex quadratic function, we show that (21a) and (21b) converge to a solution of problem (6) at an accelerated rate. We then remove this quadratic restriction on f and show that acceleration carries over to general strongly convex problems for FB splitting dynamics (21a).

3.1 Convex quadratic f

We first restrict our attention to a nonsmooth optimization problem (6) with quadratic f . In Theorem 1, we establish accelerated sublinear convergence for convex problems and, in Theorem 2, we show that the convergence rate can be improved to exponential for strongly convex problems. This extends results from the literature from a discrete-time to a continuous-time setting.

Assumption 2 *In (6), f is a convex quadratic function with a strong convexity constant $m \geq 0$ and a Lipschitz continuity constant L , and g is a proper closed convex function.*

Theorem 1 *Let Assumption 2 hold and let $\mu \in (0, 1/L)$. If the algorithmic parameters are given by $(\alpha > 0, \gamma(t) = 3/(t+3), \beta(t) = 1 - \gamma(t))$ then the solutions to (21a) and (21b) for $t \geq 0$ satisfy*

$$F(p_\mu(x(t))) - F(x^*) \leq \frac{c_1}{(t+3)^2} (\|x(0) - x^*\|_2^2 + \|\dot{x}(0)\|_2^2)$$

where $p_\mu(x) := \mathbf{prox}_{\mu g}(x - \mu \nabla f(x))$, $x^* \in \operatorname{argmin} F(x)$, and $c_1 > 0$ is a constant that depends on the problem parameters.

Theorem 1 establishes accelerated sublinear convergence rate for convex problems with quadratic smooth part of the objective function. This matches the optimal rate $O(1/k^2)$ of discrete-time gradient-based methods [14]. Furthermore, in Theorem 2 we show that the use of constant algorithmic parameters for strongly convex problems with quadratic f and known $m > 0$ leads to an accelerated exponential convergence rate. In proof of Theorem 2, we utilize Lemma 2 which provides estimates of

the smoothness and strong convexity constants of the FB and DR envelopes under Assumption 2.

Lemma 2 (Thm. 2.3 [21] and Proposition 4.6 [8]) *Let Assumption 2 hold and let $\mu \in (0, 1/L)$. The FB and DR envelopes \mathcal{F}_μ and \mathcal{D}_μ are convex functions with the following smoothness and strong convexity constants*

$$\begin{aligned} \tilde{L} &= \begin{cases} 2(1 - \mu m)/\mu & \text{for } \mathcal{F}_\mu \\ (1 - \mu m)/(\mu(1 + \mu m)^2) & \text{for } \mathcal{D}_\mu \end{cases} \\ \tilde{m} &= \begin{cases} \min\{(1 - \mu m)m, (1 - \mu L)L\} & \text{for } \mathcal{F}_\mu \\ \min\left\{\frac{(1 - \mu m)m}{(1 + \mu m)^2}, \frac{(1 - \mu L)L}{(1 + \mu L)^2}\right\} & \text{for } \mathcal{D}_\mu \end{cases} \end{aligned}$$

If the condition number $\kappa = L/m$ of f satisfies $\kappa \geq 2$, then Lemma 2 shows that for $\mu = 1/(kL)$ with $k \geq 2$, the condition numbers of the FB and DR envelopes, $\tilde{\kappa} := \tilde{L}/\tilde{m}$, are both upper bounded by $2k\kappa$.

Theorem 2 *Let Assumption 2 hold with $m > 0$, let $\mu \in (0, 1/L)$, and let \tilde{L} and \tilde{m} be the smoothness and strong convexity constants of either the FB or the DR envelope. If the algorithmic parameters in (21) are given by $(\alpha > 0, \gamma = 2\sqrt{\alpha\tilde{m}}/(\sqrt{\alpha\tilde{m}} + 1), \beta = 1 - \gamma)$ then the solutions to (21a) and (21b) for $t \geq 0$ satisfy*

$$\|x(t) - x^*\|_2^2 \leq c_2 (\|x(0) - x^*\|_2^2 + \|\dot{x}(0)\|_2^2) e^{-\rho t}$$

where $\rho = \sqrt{\alpha\tilde{m}} - \alpha\tilde{m}/2$ is the convergence rate and $c_2 > 0$ is a constant that depends on the problem parameters.

3.2 Strongly convex f

We next show that an additional restriction on μ allows us to obtain an accelerated exponential convergence rate for (21a) even when the smooth part of the objective function f in (6) is not quadratic but strongly convex.

Theorem 3 *Let Assumption 1 hold and let $\mu \in (0, \sqrt{\gamma\beta}/(2L)]$. If the algorithmic parameters in (21a) are given by $(\alpha > 0, \gamma = 2\sqrt{\alpha\mu}/(\sqrt{\alpha\mu} + 1), \beta = 1 - \gamma)$ then the solutions to (21a) for $t \geq 0$ satisfy*

$$\|x(t) - x^*\|_2^2 \leq c_3 (\|x(0) - x^*\|_2^2 + \|\dot{x}(0)\|_2^2) e^{-\rho t}$$

where $\rho = \sqrt{\alpha\mu} - \alpha\mu/2$ is the convergence rate, and $c_3 > 0$ is a constant that depends on the problem parameters.

Setting $\alpha = 1/\tilde{L}$ and $\alpha = 1/L$ in Theorems 2 and 3, respectively, gives an exponential convergence rate, $\rho \sim O(1/\sqrt{\kappa})$. This rate matches the best-achievable rate of discrete-time gradient-based methods for strongly convex problems [15]. Moreover, for $\alpha = 1/L$, the upper bound on the penalty parameter μ in Theorem 3 is of

$O(1/(L\sqrt[4]{\kappa}))$ and is smaller than the upper bound on μ in Theorem 2 by a factor of $O(1/\sqrt[4]{\kappa})$.

Remark 1 For the rate-preserving discretization proposed in [13] to be applicable to accelerated dynamics (21a) and (21b), we require $\beta + \gamma = 1$ in Theorems 1-3. For general convex problems, this additional constraint on algorithmic parameters together with the lack of convexity of envelope functions makes the analysis challenging. For (21a), we obtain acceleration by introducing additional constraints on the penalty parameter μ and two novel inequalities that substitute the strong convexity inequalities; see Lemma 3 for details.

4 Proof of main results

We next present a system-theoretic framework that can be specialized to particular problem setups addressed in Theorems 1-3. To simplify notation, we assume without loss of generality that in (21a), $\min_x \mathcal{F}_\mu(x) = 0$ and that $x^* = 0$ is a minimizer; similarly, for (21b), $\min_z \mathcal{D}_\mu(z) = 0$ and $z^* = 0$ is a minimizer.

System representation. We rewrite (21) as a feedback of a linear system with a static nonlinear block,

$$\dot{\psi} = A\psi + Bu, \quad y = C\psi, \quad u = \Delta(y) \quad (22a)$$

where $\psi = [x^T \dot{x}^T]^T$ and $\psi = [z^T \dot{z}^T]^T$ denote the respective state vectors for (21a) and (21b). On the other hand, the system matrices are given by

$$\left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array} \right] = \left[\begin{array}{cc|c} 0 & I & 0 \\ 0 & -\gamma I & -\alpha I \\ \hline I & \beta I & 0 \end{array} \right]. \quad (22b)$$

Here, the partition is done conformably with the partition of the state vector $\psi = [\psi_1^T \psi_2^T]^T$ and the corresponding nonlinear terms are

$$\Delta(y) = \begin{cases} G_\mu(y), & \text{FB splitting} \\ G_\mu(\mathbf{prox}_{\mu f}(y)), & \text{DR splitting.} \end{cases} \quad (22c)$$

4.1 Theorems 1 and 2: convex quadratic f

We first prove our results for convex quadratic f in (6), $f(x) = (1/2)x^T Qx + q^T x$, where $Q \in \mathbb{R}^{n \times n}$ is a positive semidefinite matrix and $q \in \mathbb{R}^n$. In this case, the Lipschitz continuity and strong convexity parameters L and m are given by the largest and smallest eigenvalues of Q .

Lyapunov function. Inspired by (5) which was used to establish accelerated convergence rates for (4) in [13], we propose the Lyapunov function candidate for (21),

$$V(t; \psi) = \alpha \mathcal{E}(\psi_1) + (1/2)\|\theta(t)\psi_1 + \psi_2\|_H^2 \quad (23)$$

where the envelope \mathcal{E} (either \mathcal{F}_μ or \mathcal{D}_μ) and the associated weight matrix H are determined by the selection of the nonlinear term Δ in (22c), and $\|x\|_H^2 := \langle x, x \rangle_H = x^T Hx$. Our aim is to show that $\dot{V}(t) + \theta(t)V(t) \leq 0$ for $t \geq t_0$ which in conjunction with the Grönwall inequality implies $V(t) \leq V(t_0) \exp(-\int_{t_0}^t \theta(\tau) d\tau)$. The inequalities presented in the theorems are then obtained by leveraging various properties of associated envelopes and the proximal operators. Hence, $\theta(t)$ in (23) determines the convergence rate. As we elaborate in the forthcoming subsections, in Theorem 1 we set $\theta(t) = 2/(t+3)$ and in Theorem 2 we set $\theta(t) = \rho$ to establish sublinear and exponential convergence rates, respectively.

Characterization of FB and DR envelopes. To upper bound θV , we first derive an upper bound on the envelope function \mathcal{E} . For a (strongly) convex quadratic f , the FB and DR envelopes \mathcal{E} are also (strongly) convex; see Lemma 2. Thus, for any $x, x' \in \mathbb{R}^n$,

$$\mathcal{E}(x) - \mathcal{E}(x') \leq \langle \nabla \mathcal{E}(x), x - x' \rangle - \frac{\tilde{m}}{2} \|x - x'\|_2^2 \quad (24)$$

where \tilde{m} is the parameter of strong convexity of \mathcal{E} . Moreover, (9) and (13) yield $\nabla \mathcal{E}(x) = H\Delta(x)$. For the FB envelope $\mathcal{E} = \mathcal{F}_\mu$, $\Delta(x) = G_\mu(x)$, and

$$H = I - \mu \nabla^2 f(x) = I - \mu Q.$$

On the other hand, for the DR envelope $\mathcal{E} = \mathcal{D}_\mu$, $\Delta(x) = G_\mu(\mathbf{prox}_{\mu f}(x))$ and the matrix H is determined by

$$H = 2\nabla \mathbf{prox}_{\mu f}(x) - I = (I - \mu Q)(I + \mu Q)^{-1}.$$

In both cases, $0 \prec H \preceq I$ for $\mu \in (0, 1/L)$, which together with (24) implies that for any $x, x' \in \mathbb{R}^n$

$$\mathcal{E}(x) - \mathcal{E}(x') \leq \langle \Delta(x), x - x' \rangle_H - \frac{\tilde{m}}{2} \|x - x'\|_H^2. \quad (25)$$

An upper bound on the Lyapunov function. We first rewrite the Lyapunov function candidate using system representation (22a) as

$$V(t; \psi) = \alpha \mathcal{E}(C_1 \psi) + (1/2)\|R(t)\psi\|_H^2 \quad (26)$$

where $C_1 = [I \ 0]$, $R(t) = [\theta(t)I \ I]$. We define $C_2 = [0 \ \beta I]$, $v = \Delta(C_1 \psi)$, and recall our assumption that $\psi_1^* = C_1 \psi^* = 0$ and $\mathcal{E}(\psi_1^*) = 0$. Using inequality (25) twice, for $(x, x') = (C\psi, C_1 \psi^*)$ and $(x, x') = (C_1 \psi, C\psi)$, along with $u = \Delta(C\psi)$ and $C = C_1 + C_2$ yield,

$$\begin{aligned} \mathcal{E}(C\psi) &\leq \langle u, C\psi \rangle_H - (\tilde{m}/2)\|C\psi\|_H^2 \\ \mathcal{E}(C_1 \psi) - \mathcal{E}(C\psi) &\leq -\langle v, C_2 \psi \rangle_H - (\tilde{m}/2)\|C_2 \psi\|_H^2. \end{aligned}$$

Finally, using the definition of V in (26) and summing the last two inequalities gives an upper bound on θV ,

$$\begin{aligned} \theta V &\leq \alpha\theta (\langle u, C\psi \rangle_H - \langle v, C_2\psi \rangle_H) \\ &\quad - \frac{\alpha\tilde{m}\theta}{2} (\|C\psi\|_H^2 + \|C_2\psi\|_H^2) + \frac{\theta}{2}\|R\psi\|_H^2. \end{aligned} \quad (27)$$

We next examine the derivative of V along the solutions of (22) and use (27) to obtain an upper bound on $\dot{V} + \theta V$.

Lyapunov analysis. The derivative of (26) along the solutions of (22) is determined by

$$\dot{V} = \alpha\langle v, C_1(A\psi + Bu) \rangle_H + \langle R\psi, \dot{R}\psi + R(A\psi + Bu) \rangle_H \quad (28)$$

where $\dot{R} = [(d\theta(t)/dt)I \ 0]$. Finally, combining (28) with (27) leads to

$$\dot{V} + \theta V \leq \frac{1}{2} \begin{bmatrix} \psi \\ u \\ v \end{bmatrix}^T \begin{bmatrix} \Pi & \Gamma & \Lambda \\ \Gamma^T & 0 & 0 \\ \Lambda^T & 0 & 0 \end{bmatrix} \begin{bmatrix} \psi \\ u \\ v \end{bmatrix} \quad (29)$$

where matrices Π , Γ , and Λ differ for convex and strongly convex cases. Now, we prove that the matrix on the right-hand-side of (29) is negative semidefinite for the parameters given in Theorems 1 and 2.

4.1.1 Theorem 1: convex quadratic f

For a (non-strongly) convex quadratic function f , $\tilde{m} = 0$, the upper bound on θV given in (27) simplifies to

$$\theta V \leq \alpha\theta (\langle u, C\psi \rangle_H + \langle v, -C_2\psi \rangle_H) + (\theta/2)\|R\psi\|_H^2.$$

We let $P := R^T H R$ and utilize this upper bound and (28) to write Π in (29) as

$$\begin{aligned} \Pi &= A^T P + P A + \theta P + \dot{P} \\ &= \begin{bmatrix} 2\dot{\theta}\theta + \theta^3 & \dot{\theta} + 2\theta^2 - \theta\gamma \\ \dot{\theta} + 2\theta^2 - \theta\gamma & 3\theta - 2\gamma \end{bmatrix} \otimes H \end{aligned} \quad (30)$$

where $\dot{P} = \dot{R}^T H R + R^T H \dot{R}$ and \otimes is the Kronecker product. A necessary condition for negative semidefiniteness of the matrix on the right-hand-side of (29) is the negative semidefiniteness of Π . One way to simplify the conditions that guarantee $\Pi \preceq 0$ is to set the off-diagonal blocks in (30) to zero, i.e.,

$$\dot{\theta} + 2\theta^2 - \theta\gamma = 0. \quad (31)$$

When (31) is satisfied, $\Pi \preceq 0$ if its diagonal blocks are negative semidefinite. The first diagonal block requires $\theta \leq 2/(t+r)$ and the largest θ that satisfies this condition is $\theta(t) = 2/(t+r)$ for some $r > 0$. Substituting this value to (31) yields $\gamma(t) = 3/(t+r)$. For this choice of

parameters, $\Pi \equiv 0$. While any choice of $r > 0$ works here, different constants c_1 in Theorem 1 are obtained for different values of r . In what follows, we set $r = 3$ to ensure positivity of $\beta(t)$ in Theorem 1 for any $t \geq 0$.

Since the off-diagonal blocks in (29) are given by

$$\begin{aligned} \Gamma &= P B + \alpha\theta C^T H = \begin{bmatrix} 0 \\ -\alpha(1 - \theta\beta)H \end{bmatrix} \\ \Lambda &= \frac{\alpha}{\beta}(1 - \theta\beta)C_2^T H = \begin{bmatrix} 0 \\ \alpha(1 - \theta\beta)H \end{bmatrix} = -\Gamma. \end{aligned} \quad (32)$$

for the above selection of θ and γ , (29) simplifies to

$$\dot{V} + \theta V \leq -(\alpha/\beta)(1 - \theta\beta)\langle u - v, C_2\psi \rangle_H. \quad (33)$$

Moreover, monotonicity of the gradient of the envelope function implies $\langle u - v, C_2\psi \rangle_H \geq 0$. Thus, the right-hand-side of (33) is upper bounded by zero and the use of Grönwall inequality yields $V(t) \leq V(0)/(t+3)^2$, $t \geq 0$. The inequality in Theorem 1 is obtained as follows. We first use the smoothness of the envelope function together with our assumption that $\min_x \mathcal{F}_\mu(x) = 0$ and $x^* \in \operatorname{argmin}_x \mathcal{F}_\mu(x)$ to obtain

$$\mathcal{E}(\psi_1(0)) \leq \frac{\tilde{L}}{2}\|\psi_1(0)\|^2 \leq \frac{\tilde{L}}{2}\|\psi(0)\|^2$$

where \tilde{L} is given in Lemma 2. Then, application of the triangle inequality, $H \preceq I$, and the fact that $\theta(t) \leq 1$ for $t \geq 0$ yield the upper bound on the Lyapunov function,

$$\begin{aligned} V(\psi(0)) &= \frac{\alpha\tilde{L}}{2}\|\psi(0)\|^2 + \frac{1}{2}\|\theta(t)\psi_1(0) + \psi_2(0)\|_H^2 \\ &\leq (\frac{\alpha\tilde{L}}{2} + 1)\|\psi(0)\|^2. \end{aligned} \quad (34)$$

We note that for both (21a) and (21b), $\mathcal{E}(\psi_1) = \mathcal{F}_\mu(x)$. Hence, a lower bound on the Lyapunov function can be obtained using the identity [21, Theorem 2.2], $F(p_\mu(x)) \leq \mathcal{F}_\mu(x)$ for any $x \in \mathbb{R}^n$. This lower bound in conjunction with (34) leads to the inequality in Theorem 1 for (21a) with constant $c_1 = 1/\tilde{L} + 1/\alpha$.

In the case of (21b) for which $\psi = [z^T \ \dot{z}^T]^T$, while the lower bound on the Lyapunov function stays the same, the upper bound (34) can be expressed in terms of x and \dot{x} using the following relations [20],

$$\begin{aligned} x &= \mathbf{prox}_{\mu f}(z) = (I + \mu Q)^{-1}z, \\ \dot{x} &= \nabla \mathbf{prox}_{\mu f}(z)\dot{z} = (I + \mu Q)^{-1}\dot{z} \end{aligned}$$

which lead to the inequality in Theorem 1 with constant $c_1 = (1 + \mu L)(1/\tilde{L} + 1/\alpha)$.

4.1.2 Theorem 2: strongly convex quadratic f

For a strongly convex f , the use of constant α , β , and γ in (21) leads to time-invariant dynamics (22a). In this case, the Lyapunov function candidate does not explicitly depend on time, i.e., R in (26) becomes a constant matrix. We again let $P = R^T H R$ and utilize (27) and (28) to write Π in (29) as

$$\Pi = A^T P + P A + \theta P - \alpha \tilde{m} \theta (C^T H C + C_2^T H C_2).$$

Since the off-diagonal blocks Γ and Λ in (29) remain the same as in (32), we have

$$\langle \psi, \Gamma u + \Lambda v \rangle = -(\alpha/\beta)(1 - \theta\beta) \langle C_2 \psi, u - v \rangle_H. \quad (35)$$

Moreover, strong monotonicity of gradient of the envelope function along with $H \preceq I$ implies $\langle u - v, C_2 \psi \rangle_H \leq \tilde{m} \|C_2 \psi\|_H^2$ which together with (35) simplifies (29) to

$$\dot{V} + \theta V \leq (1/2) \psi^T (\Pi - \Upsilon) \psi \quad (36)$$

where

$$\Upsilon = (2\alpha\tilde{m}/\beta)(1 - \theta\beta) C_2^T H C_2.$$

Substitution of the system matrices in (22b) gives

$$\Pi - \Upsilon = \begin{bmatrix} \theta(\theta^2 - \alpha\tilde{m}) & \theta(2\theta - \gamma - \alpha\tilde{m}\beta) \\ \theta(2\theta - \gamma - \alpha\tilde{m}\beta) & 3\theta - 2\gamma - 2\alpha\tilde{m}\beta \end{bmatrix} \otimes H.$$

The conditions that guarantee $\Pi - \Upsilon \preceq 0$ can be simplified by setting off-diagonal blocks in $\Pi - \Upsilon$ to zero, i.e.,

$$\theta = (\gamma + \alpha\tilde{m}\beta)/2. \quad (37)$$

When (37) is satisfied, the second diagonal block in $\Pi - \Upsilon$ is clearly negative definite. Since $\beta = 1 - \gamma$, the negative semidefiniteness of the first diagonal block requires $\gamma \leq (2\sqrt{\alpha\tilde{m}} - \alpha\tilde{m})/(1 - \alpha\tilde{m})$ which is trivially satisfied by the choice of $\gamma = (2\sqrt{\alpha\tilde{m}} - 2\alpha\tilde{m})/(1 - \alpha\tilde{m})$ given in the theorem. Hence, the right-hand-side of (36) is negative semidefinite, leading to $V(t) \leq V(0)e^{-\theta t}$ for $t \geq 0$.

The strong convexity and smoothness of the envelope functions yield quadratic lower and upper bounds on the Lyapunov function, which completes the proof for (21a). For (21b), the proof is completed as in Section 4.1.1, i.e., utilizing linearity of the proximal operator together with $\dot{x} = \nabla \text{prox}_{\mu f}(z)\dot{z}$ and (12).

4.2 Theorem 3: strongly convex f

To prove our results for general strongly convex f , we use system representation given by (22) with constant parameters but with a different Lyapunov function and a different characterization of the FB envelope.

Lyapunov function. For general strongly convex f , since the weight matrix $H(t) = I - \mu \nabla^2 f(x(t))$ is time-dependent, in lieu of (26) we propose the following time-invariant Lyapunov function candidate

$$V(\psi) = \alpha \mathcal{F}_\mu(C\psi) + (1/2) \psi^T P \psi \quad (38)$$

where C is given in (22b), $P = R^T R$, and R is given in (26). In contrast to (26), the envelope function in (38) is evaluated at $y = C\psi$ and the quadratic form does not use the time-dependent weight matrix H .

Characterization of the FB envelope. For general strongly convex problems, the FB envelope is not necessarily convex and the upper bounds (24) and (25) no longer hold. In Lemmas 3, we derive two novel inequalities that substitute (24); see Appendix for the proof.

Lemma 3 *Let Assumption 1 hold and let $\mu \in (0, 1/L)$. For every $x, x' \in \mathbb{R}^n$ and $x^* \in \text{argmin} F(x)$, FB envelope satisfies*

$$\mathcal{F}_\mu(x) - F(x') \leq \langle G_\mu(x), x - x' \rangle - \frac{m}{2} \|x - x'\|_2^2 - \frac{\mu}{2} \|G_\mu(x)\|_2^2. \quad (39a)$$

$$\mathcal{F}_\mu(x) - F(x^*) \geq \frac{m^2(1 - \mu L)}{2L} \|x - x^*\|_2^2. \quad (39b)$$

We note that in contrast to strong convexity inequality (24), in Lemma 3, the gradient of \mathcal{F}_μ is not involved and we characterize the difference between envelope \mathcal{F}_μ and objective function F . Still, Lyapunov function (38) can be upper bounded via Lemma 2 based on the equivalence given in (14).

An upper bound on the Lyapunov function. The following upper bound on (38) can be directly obtained using Lemma 3, (14), and system representation (22a),

$$\theta V \leq \alpha \theta \langle u, C\psi \rangle - \frac{\alpha m \theta}{2} \|C\psi\|_2^2 - \frac{\alpha \mu \theta}{2} \|u\|_2^2 + \frac{\theta}{2} \psi^T P \psi. \quad (40)$$

Lyapunov analysis. The derivative of (38) along the solutions of (21a) is determined by

$$\dot{V} = \alpha \langle H u, C A \psi + C B u \rangle + \langle P \psi, A \psi + B u \rangle.$$

Summing \dot{V} with the upper bound given in (40) yields

$$\dot{V} + \theta V \leq \frac{1}{2} \begin{bmatrix} \psi \\ u \end{bmatrix}^T \begin{bmatrix} \Pi & \Gamma + \Lambda H \\ \Gamma^T + H \Lambda^T & -2\alpha^2 \beta H \end{bmatrix} \begin{bmatrix} \psi \\ u \end{bmatrix}$$

where

$$\begin{aligned}\Pi &= A^T P + PA + \theta P - \alpha m \theta C^T C \\ \Gamma &= PB + \alpha \theta C^T \\ \Lambda &= \alpha A^T C^T.\end{aligned}\quad (41)$$

Since $H \succ 0$, the Schur complement implies that $\dot{V} + \theta V \leq 0$ if and only if

$$2\alpha^2 \beta \Pi + (\Gamma \Lambda^T + \Lambda \Gamma^T) + \Gamma H^{-1} \Gamma^T + \Lambda H \Lambda^T \preceq 0. \quad (42)$$

Substitution of system matrices (22b) into (41) gives

$$\Gamma \Lambda^T + \Lambda \Gamma^T = \begin{bmatrix} 0 & 0 \\ 0 & -2\alpha^2(1 - \theta\beta)(1 - \gamma\beta)I \end{bmatrix} \quad (43a)$$

$$\Gamma H^{-1} \Gamma^T + \Lambda H \Lambda^T = \begin{bmatrix} 0 & 0 \\ 0 & \alpha^2((1 - \theta\beta)^2 H^{-1} + (1 - \gamma\beta)^2 H) \end{bmatrix}. \quad (43b)$$

The eigenvalue decomposition of H can be used to show

$$\Gamma H^{-1} \Gamma^T + \Lambda H \Lambda^T \preceq \begin{bmatrix} 0 & 0 \\ 0 & \alpha^2 \eta I \end{bmatrix} \quad (44a)$$

$$\eta := \max_{\sigma \in [m, L]} \frac{(1 - \theta\beta)^2}{1 - \mu\sigma} + (1 - \gamma\beta)^2(1 - \mu\sigma). \quad (44b)$$

If $(1 - \theta\beta) \geq (1 - \mu\sigma)(1 - \gamma\beta)$ for all $\sigma \in [m, L]$, then the objective function in (44b) is increasing for $\sigma \in [m, L]$ and the maximum occurs at $\sigma = L$,

$$\eta = (1 - \theta\beta)^2/(1 - \mu L) + (1 - \gamma\beta)^2(1 - \mu L). \quad (45)$$

Combining (43a) with (44a) and (45) yields

$$\begin{aligned} & (\Gamma \Lambda^T + \Lambda \Gamma^T) + \Gamma H^{-1} \Gamma^T + \Lambda H \Lambda^T \preceq \\ & \begin{bmatrix} 0 & 0 \\ 0 & \frac{\alpha^2}{1 - \mu L} ((1 - \theta\beta) - (1 - \mu L)(1 - \gamma\beta))^2 I \end{bmatrix}. \end{aligned}$$

Moreover, since the entries of the matrix Π are given by

$$\Pi = \begin{bmatrix} \theta(\theta^2 - \alpha m) & \theta(2\theta - \gamma - \alpha m\beta) \\ \theta(2\theta - \gamma - \alpha m\beta) & 3\theta - 2\gamma - \alpha m\theta\beta^2 \end{bmatrix} \otimes I$$

setting the off-diagonal blocks in Π to zero yields

$$\theta = (\gamma + \alpha m\beta)/2. \quad (46)$$

When (46) is satisfied, verifying $\dot{V} + \theta V \leq 0$, or equivalently (42), amounts to checking these three conditions,

- (i) $(1 - \theta\beta) \geq (1 - \mu\sigma)(1 - \gamma\beta)$ for all $\sigma \in [m, L]$
- (ii) $\theta^2 \leq \alpha m$
- (iii) $\frac{((1 - \theta\beta) - (1 - \mu L)(1 - \gamma\beta))^2}{2\beta(1 - \mu L)} \leq \alpha m\theta\beta^2 + 2\gamma - 3\theta$.

While condition (i) is necessary for upper bound (45), conditions (ii) and (iii) ensure negative semidefiniteness of the diagonal blocks in (42).

For γ given in Theorem 3 and $\beta = 1 - \gamma$, we obtain $\theta = \sqrt{\alpha m} - \alpha m/2$ using (46). Thus, $\theta \leq \gamma$ and conditions (i) and (ii) are trivially satisfied. For large values of $\kappa = L/m$, the numerator of the left-hand-side in condition (iii) is dominated by μL . On the other hand, the largest term in the product of the denominator and the right-hand-side is $\gamma\beta$, which suggests that condition (iii) holds if $\mu \sim O(\sqrt{\gamma\beta}/L)$. Due to space limitations, we use computational tools to verify this hypothesis. With the above choices of (γ, β, θ) and $\mu L = \sqrt{\gamma\beta}/2$, condition (iii) can be written as $wh(w) \leq 0$ where $w = \sqrt{\alpha m} \in [0, 1]$ and $h: [0, 1] \rightarrow \mathbb{R}$. In Fig. 2(a), we show that h is negative for all $w \in [0, 1]$. Moreover, since (i) implies that $wh(w)$ is an increasing function of μL , condition (iii) is satisfied for all $\mu L \in (0, \sqrt{\gamma\beta}/2]$.

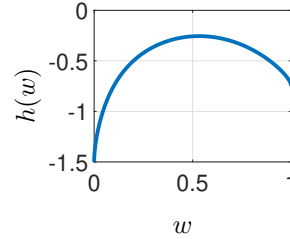


Fig. 1. Plot of the function h that demonstrates $h < 0$. Thus, $wh(w) \leq 0$ for $w \in [0, 1]$ and condition (iii) is satisfied.

As a result, using Grönwall inequality, we conclude that $V(t) \leq V(0)e^{-\theta t}$ for $t \geq 0$ along the solutions of (21a). A quadratic upper bound on the Lyapunov function can be obtained by using the smoothness of the FB envelope as in the proof of Theorem 1. Moreover, since $C^T C + P$ is invertible, inequality (39b) in Lemma 3 yields a quadratic lower bound on $V(t)$. These upper and lower quadratic bounds complete the proof.

5 Computational experiments

We use Matlab's function ode45 to evaluate performance of our approach. We choose zero initial conditions, $\alpha = 1/L$, and provide comparison with Accelerated ADMM and Tseng's Splitting [7]. Parameters for these discrete-time algorithms are selected as described in [7, Sec. IX.C] with decaying (Example 1) and constant (Example 3) damping. Error trajectories of these are plotted at times determined by $t_k = kh/\alpha$, where k is the iteration index, $h = 1/L$ is the stepsize of the discrete-time algorithm, and α is the time scale in (21). While all accelerated algorithms exhibit similar convergence behavior in the convex case (Example 1), our proposed accelerated

proximal gradient flow achieves slightly faster convergence in the strongly convex problem (Example 3).

5.1 Example 1: ℓ_1 -regularized least squares

Let us examine an ℓ_1 -regularized least squares problem,

$$\underset{x}{\text{minimize}} \quad (1/2)\|Ex - q\|_2^2 + \lambda\|x\|_1 \quad (47)$$

where the problem data $E \in \mathbb{R}^{100 \times 2000}$, $q \in \mathbb{R}^{100}$, $\lambda > 0$ are generated as described in [18]. Since $E^T E$ is a singular matrix, this problem is not strongly convex. Thus, the time-varying parameters given in Theorem 1 with $\mu = 1/(2L)$ are used to illustrate improved sublinear convergence rate of accelerated proximal gradient flow and DR splitting dynamics (21a) and (21b) in Fig. 2(b).

5.2 Example 2: Box-constrained quadratic program

For a box-constrained quadratic program,

$$\underset{x}{\text{minimize}} \quad (1/2)x^T Qx + q^T x + I_{\mathcal{C}}(x) \quad (48)$$

where $I_{\mathcal{C}}(x)$ is the indicator function of the convex set $\mathcal{C} = \{x \mid l \leq x \leq u\}$, we generate problem data with positive definite $Q \in \mathbb{R}^{500 \times 500}$, and vectors $q, l, u \in \mathbb{R}^{500}$ as described in [9] and set the condition number $\kappa = 10^5$. Linear convergence of the algorithms for constant parameters given in Theorem 2 is illustrated in Fig. 2(b).

5.3 Example 3: ℓ_1 -regularized logistic regression

We examine the ℓ_1 -regularized logistic regression problem

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^n \left(-y_i a_i^T x + \log(1 + e^{a_i^T x}) \right) + \lambda\|x\|_1 + \frac{\rho}{2}\|x\|_2^2 \quad (49)$$

where the problem data $A = [a_1 \ \dots \ a_n]^T \in \mathbb{R}^{200 \times 1000}$ as well as the scalars λ and y_i are generated as described in [24]. In Fig. 2(c), we validate linear convergence of accelerated proximal gradient flow dynamics (21a) with constant parameters given in Theorem 3 and $\mu = 1/(L\sqrt[4]{\kappa}) \approx 1/(26L)$ for this strongly convex problem with $\rho = 0.1$, and the condition number $\kappa \approx 5 \cdot 10^5$.

6 Concluding remarks

We have introduced continuous-time accelerated FB and DR splitting dynamics for composite optimization problems where the objective function is given by the sum of smooth and nonsmooth terms. Motivated by the variable-metric gradient method interpretation of FB and DR splittings, we have leveraged smoothness of associated envelopes and utilized a Lyapunov-based approach to establish convergence rates for both convex and strongly convex problems.

A Proof of Lemma 3

Our proof is based on an equivalent definition of the FB envelope [21, Eq. (2.8b)]

$$\mathcal{F}_\mu(x) = f(x) + g(p(x)) - \mu \langle \nabla f(x), G_\mu(x) \rangle + \frac{\mu}{2} \|G_\mu(x)\|_2^2 \quad (\text{A.1})$$

and the subgradient inequality,

$$g(x) \geq g(x') + \langle r, x - x' \rangle, \quad \forall x, x' \in \mathbb{R}^n, \quad \forall r \in \partial g(x'). \quad (\text{A.2})$$

Proof of (39a). Let $p(x) := \mathbf{prox}_{\mu g}(x - \mu \nabla f(x))$. The identity (8) together with (10) implies that

$$G_\mu(x) - \nabla f(x) \in \partial g(p(x)).$$

Hence, the subgradient inequality (A.2) for an arbitrary $x' \in \mathbb{R}^n$ yields

$$g(x') \geq g(p(x)) + \langle G(x) - \nabla f(x), x' - p(x) \rangle$$

which in conjunction with (A.1) results in

$$\begin{aligned} \mathcal{F}_\mu(x) &\leq f(x) + g(x') - \langle G(x) - \nabla f(x), x' - p(x) \rangle \\ &\quad - \mu \langle \nabla f(x), G_\mu(x) \rangle + \frac{\mu}{2} \|G_\mu(x)\|_2^2. \end{aligned}$$

Rearranging terms and then upper bounding $f(x)$ via strong convexity,

$$f(x) \leq f(x') + \langle \nabla f(x), x - x' \rangle - \frac{m}{2} \|x - x'\|^2$$

gives

$$\begin{aligned} \mathcal{F}_\mu(x) &\leq f(x') + g(x') - \frac{m}{2} \|x - x'\|^2 \\ &\quad + \langle G(x), p(x) - x' \rangle + \frac{\mu}{2} \|G_\mu(x)\|_2^2 \\ &= F(x') - \frac{m}{2} \|x - x'\|^2 + \langle G_\mu(x), x - x' \rangle \\ &\quad - \langle G_\mu(x), x - p(x) \rangle + \frac{\mu}{2} \|G_\mu(x)\|_2^2 \end{aligned}$$

where the equality is obtained by adding and subtracting x from the second argument of the inner product. Substituting (10) into the second inner product completes the proof.

Proof of (39b). Let $p(x) := \mathbf{prox}_{\mu g}(x - \mu \nabla f(x))$, $\tilde{x} := x - x^*$, and $\tilde{p} := p(x) - p(x^*)$. Substitution of (10) into (A.1) gives

$$\begin{aligned} \mathcal{F}_\mu(x) - F(x^*) &= f(x) - f(x^*) + g(p(x)) - g(x^*) \\ &\quad - \langle \nabla f(x), \tilde{x} - \tilde{p} \rangle + \frac{\mu}{2} \|G_\mu(x)\|_2^2. \end{aligned} \quad (\text{A.3})$$

The optimality condition for problem (6) implies that $-\nabla f(x^*) \in \partial g(x^*)$ and $x^* = p(x^*)$. Using these two identities in the subgradient inequality (A.2) yields

$$g(p(x)) \geq g(x^*) - \langle \nabla f(x^*), \tilde{p} \rangle$$

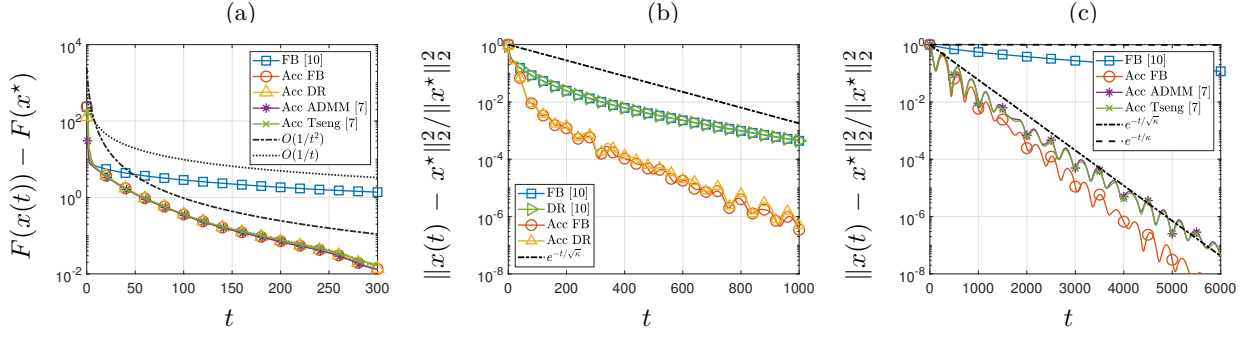


Fig. 2. (a) Sublinear convergence of accelerated dynamics (21a) and (21b) with the time-varying parameters given in Theorem 1 for ℓ_1 -regularized least squares problem (47). (b) Linear convergence of accelerated dynamics (21a) and (21b) with the constant parameters given in Theorem 2 for box-constrained quadratic program (48). (c) Linear convergence of accelerated dynamics (21a) with the constant parameters given in Theorem 3 and $\mu = 1/(L\sqrt[4]{\kappa}) \approx 1/(26L)$ for the ℓ_1 -regularized logistic regression problem (49) with $\varrho = 0.1$, and condition number $\kappa \approx 5 \cdot 10^5$.

which together with (A.3) results in

$$\mathcal{F}_\mu(x) - F(x^*) \geq f(x) - f(x^*) - \langle \nabla f(x), \tilde{x} \rangle + \langle \nabla f(x) - \nabla f(x^*), \tilde{p} \rangle + \frac{\mu}{2} \|G_\mu(x)\|^2.$$

Substitution of the lower bound given by the cocoercivity of gradient $\nabla f(x)$,

$$f(x) - f(x^*) \geq \langle \nabla f(x^*), \tilde{x} \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(x^*)\|^2$$

into the right-hand-side yields

$$\begin{aligned} \mathcal{F}_\mu(x) - F(x^*) &\geq \langle \nabla f(x^*) - \nabla f(x), \tilde{x} - \tilde{p} \rangle \\ &\quad + \frac{1}{2L} \|\nabla f(x) - \nabla f(x^*)\|^2 + \frac{\mu}{2} \|G_\mu(x)\|^2 \\ &= \frac{\mu}{2} \|G_\mu(x) - (\nabla f(x) - \nabla f(x^*))\|^2 \\ &\quad + \frac{1}{2L} (1 - \mu L) \|\nabla f(x) - \nabla f(x^*)\|^2 \\ &\geq \frac{m^2}{2L} (1 - \mu L) \|x - x^*\|^2 \end{aligned}$$

where the equality follows from (10) and the completion of squares; the second inequality is obtained by ignoring the first (nonnegative) term and using m -strong convexity of f on the second one.

References

- [1] A. Allibhoy and J. Cortés. Control barrier function-based design of gradient flows for constrained nonlinear programming. *IEEE Trans. Automat. Control*, 2023.
- [2] H. Attouch and Z. Chbani. Fast inertial dynamics and FISTA algorithms in convex optimization. Perturbation aspects. *arXiv preprint arXiv:1507.01367*, 2015.
- [3] H. Attouch, Z. Chbani, J. Fadili, and H. Riahi. Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics. *J. Optim. Theory Appl.*, 193(1-3):704–736, 2022.
- [4] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.*, 168:123–175, 2018.
- [5] P. Cisneros-Velarde and F. Bullo. A contraction theory approach to optimization algorithms from acceleration flows. In *Proc. Int. Conf. Artif. Intell. Stat.*, pages 1321–1335, 2022.
- [6] N. K. Dhingra, S. Z. Khong, and M. R. Jovanović. The proximal augmented Lagrangian method for nonsmooth composite optimization. *IEEE Trans. Automat. Control*, 64(7):2861–2868, 2019.
- [7] G. Franca, D. P. Robinson, and R. Vidal. Gradient flows and proximal splitting methods: A unified view on accelerated and stochastic optimization. *Phys. Rev. E*, 103(5):53304, 2021.
- [8] P. Giselsson and M. Falt. Envelope functions: Unifications and further properties. *J. Optim. Theory Appl.*, 178:673–698, 2018.
- [9] C. C. Gonzaga, E. W. Karas, and D. R. Rossetto. An optimal algorithm for constrained differentiable convex optimization. *SIAM J. Control Optim.*, 23(4):1939–1955, 2013.
- [10] S. Hassan-Moghaddam and M. R. Jovanović. Proximal gradient flow and Douglas-Rachford splitting dynamics: Global exponential stability via integral quadratic constraints. *Automatica J. IFAC*, 123:109311, 2021.
- [11] X. He, R. Hu, and Y. Fang. “Second-order primal”+“first-order dual” dynamical systems with time scaling for linear equality constrained convex optimization problems. *IEEE Trans. Automat. Control*, 67(8):4377–4383, 2022.
- [12] T. Liu and T. K. Pong. Further properties of the forward–backward envelope with applications to difference-of-convex programming. *Comput. Optim. Appl.*, 67:489–520, 2017.
- [13] M. Muehlebach and M. I. Jordan. A dynamical systems perspective on Nesterov acceleration. In *Proc. Int. Conf. Mach. Learn.*, pages 4656–4662, 2019.
- [14] Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In *Dokl. Akad. Nauk*, volume 269, pages 543–547, 1983.
- [15] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer, Berlin, 2003.
- [16] I. K. Ozaslan and M. R. Jovanović. Exponential convergence of primal-dual dynamics for multi-block problems under local error bound condition. In *Proc. IEEE Conf. Decis. Control*, pages 7579–7584, 2022.
- [17] I. K. Ozaslan and M. R. Jovanovic. Tight lower bounds on the convergence rate of primal-dual dynamics for equality constrained convex problems. In *Proc. IEEE Conf. Decis. Control*, pages 7318–7323, 2023.

- [18] B. O’Donoghue and E. Candes. Adaptive restart for accelerated gradient schemes. *Found. Comput. Math.*, 15:715–732, 2015.
- [19] N. Parikh and S. Boyd. Proximal algorithms. *Found. Trends Opt.*, 1(3):123–231, 2014.
- [20] P. Patrinos, L. Stella, and A. Bemporad. Douglas-Rachford splitting: Complexity estimates and accelerated variants. In *Proc. IEEE Conf. Decision Control*, pages 4234–4239, 2014.
- [21] P. Patrinos, L. Stella, and A. Bemporad. Forward-backward truncated Newton methods for convex composite optimization. *arXiv:1402.6655*, 2014.
- [22] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *Comput. Math. Phys.*, 4(5):1–17, 1964.
- [23] E. K Ryu and S. Boyd. Primer on monotone operator methods. *Appl. comput. math*, 15(1):3–43, 2016.
- [24] W. Su, S. Boyd, and E. Candes. A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *J. Mach. Learn. Res.*, 17(153):1–43, 2016.
- [25] A. Wibisono, A. C. Wilson, and M. I. Jordan. A variational perspective on accelerated methods in optimization. *Proc. Natl. Acad. Sci. USA*, 113(47):E7351–E7358, 2016.
- [26] A. C Wilson, B. Recht, and M. I. Jordan. A Lyapunov analysis of accelerated methods in optimization. *J. Mach. Learn. Res.*, 22(1):5040–5073, 2021.
- [27] X. Zeng, J. Lei, and J. Chen. Dynamical primal-dual Nesterov accelerated method and its application to network optimization. *IEEE Trans. Automat. Control*, 68(3):1760–1767, 2023.